VARIATION AND STANDARDISATION

THE CASE OF AFRIKAANS (1880-1922)

Andrea Deumert

Thesis Presented for the Degree of
Doctor of Philosophy
in the Department of Linguistics and Southern African Languages
University of Cape Town
June 1999
The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.
ABSTRACT

Following the general model outlined in Weinreich, Labov & Herzog (1968), this study is a contribution to the historiography of Afrikaans from a variationist perspective, investigating the patterns of linguistic variability in the context of the early standardisation of Afrikaans. The work is based on a newly collected historical corpus of private documents which includes letters and diaries from 136 individuals (written between 1880 and 1922), and can be said to represent acrolectal and mesolectal usage. Several morphosyntactic, morphological and syntactic variables were investigated: loss of person and number distinctions in the present tense paradigm, loss of the infinitive, regularisation of the past participle, loss of the preterite, loss of gender, the emergence of a new system of adjective inflection and of a new pronoun system, the so-called 'double' negation, infinitive clauses, the use of objective vir, and the periphrastic possessive with se. The quantitative analysis of these variables makes use of a variety of methods: descriptive techniques such as distribution analysis, implicational scaling, cluster analysis, multidimensional scaling and principal components analysis, as well as inferential statistics such as the chi-square test. Variation is furthermore described from a code-switching perspective. As a study in historical sociolinguistics this dissertation is also concerned with the epistemological aspects of sociohistorical research, in particular the role of speaker agency in historical explanations, the 'measurement' of the extralinguistic variables in sociolinguistic research, the nature of the relationship between sociolinguistic and social theory, and in general the ontological status of our explanatory and descriptive concepts and taxonomies.

While traditionally historiographers of Afrikaans have argued that there existed a sharp linguistic and functional distinction between Afrikaans and Dutch from the mid 18th century, the patterns of variability described for the corpus indicate the existence of a complex dialect continuum (rather than diglossia) until the early 20th century. The results of the quantitative analysis suggest furthermore that the process of linguistic change was slower than hitherto assumed, and variation patterns described for the late 18th century were still found to exist in the corpus. Such continuities challenge the conventional dating of the emergence of Afrikaans as a new language or dialect (characterised by almost complete morphological regularisation and a cluster of innovative syntactic features) to around 1800.

As regards the standardisation of Afrikaans the study shows that from the 1850s a relatively uniform model of what constituted 'the vernacular' (or 'Afrikaans' as it came to be known) existed as a well-defined entity in the popular consciousness, while the actual language use of many speakers remained rather more variable. Increasingly, linguistic practices which were not in line with the propagated model of 'Afrikaans' were identified by the contemporary metalinguistic discourse (which was strongly marked by 19th century cultural nationalism) as unauthentic and thus undesirable. The diffusion of the new standard is shown to have followed the path typical for modern standard languages, i.e. via the socially (and geographically) mobile professional class or intelligentsia. After about 1914 the new standard was widely diffused, and had replaced other Netherlandic varieties in many private documents.
CONTENTS

List of Figures, Maps and Tables vii

Abbreviations xiii

Acknowledgements xiv

PROLOGUE 1

PART I THEORETICAL AND SOCIOHISTORICAL BACKGROUND

1 HISTORICAL SOCIOLINGUISTICS: SOME THEORETICAL REFLECTIONS 8
  1.1 Variation Studies, Speakers and Linguistic Evolution
  1.2 The Agents of Change 14
    1.2.1 Desperately Seeking: A Theory of Action
    1.2.2 The Locus of Action: Collectives vs Individuals 22
    1.2.3 Structure, Social Praxis and the Unintended Consequences of Action 28
  1.3 Directions for Historical Sociolinguistics 37

2 AFRIKAANS HISTORICAL LINGUISTICS: MAIN POSITIONS AND DATA BASIS 40
  2.1 The Emergence of Afrikaans: Prototypes and Variability
  2.2 The Nature and Results of Language Contact at the Cape 47
    2.2.1 Sociohistorical and Linguistic Evidence 1652-1800
    2.2.2 Creolisation or Interlectal Restructuring - The Question of Stability 63
  2.3 The Insights of the Acrolect 69

3 THE 19TH CENTURY AND THE 'DISCOVERY' OF THE VERNACULAR 72
  3.1 The Rise of the Vernacular Writing Tradition
  3.2 Afrikaner Nationalism and Early Vernacular Standardisation 81
  3.3 Some Comments on Folk Taxonomies and Language Attitudes 87
  3.4 A Diglossic Community? 90
  3.5 The 'Linguistic Marketplace' and its Entrepreneurs 99
PART II DATA COLLECTION AND STATISTICAL METHODOLOGY

4 THE SURVEY: SOCIOLINGUISTIC VARIATION IN THE CAPE COLONY (1880-1922)
   4.1 Constructing the Sociohistorical Sample
      4.1.1 The Netherlandic Speech Community 104
      4.1.2 Collecting the Data 110
   4.2 The Linguistic Variables 119
   4.3 The Social (Extra-linguistic) Variables - Mapping the Social Universe
      4.3.1 Age, Sex and Ethnicity 123
      4.3.2 Social Class 130
   4.4 Summary 140

5 QUANTIFYING AND CLASSIFYING VARIATION
   5.1 Statistics and Linguistics: More than a Numbers Game 142
   5.2 Multivariate Clustering Techniques
      5.2.1 Numerical Taxonomy 145
      5.2.2 Cluster Analysis 146
      5.2.3 Multidimensional Scaling 158
      5.2.4 Principal Components Analysis 161
   5.3 Summary and Perspectives 169

PART III DESCRIPTION AND INTERPRETATION

6 MORPHOSYNTACTIC VARIATION
   6.1 The Verbal System
      6.1.1 Loss of Person and Number Distinction in the Present Tense Paradigm 172
      6.1.2 Re-visiting Conradie (1979) 183
   6.2 Tense
      6.2.1 Past Tense Variation 194
      6.2.2 Tense and Aspect Adverbials 198
   6.3 Gender
      6.3.1 Loss of Gender Agreement 202
      6.3.2 Attributive Adjective Inflection 207
   6.4 The Gradualness of Morphological Change 213
7 PATTERNS OF MORPHOLEXICAL AND SYNTACTIC VARIATION

7.1 The Pronoun System
  7.1.1 Subject and Object Pronouns 215
  7.1.2 Attributive Possessive Pronouns 222
  7.1.3 The Relative Pronoun 224
  7.1.4 The Demonstrative Pronouns *hierdie* and *daardie* 227

7.2 Some Other Morpholexical Variables (*het*, *baie* and the complementizer *lat/laat*) 230

7.3 Selected Syntactic Variables
  7.3.1 The Double or Embracing Negation 232
  7.3.2 The Infinitive with *om...te* 240
  7.3.3 Objective *Vir* 243
  7.3.4 The Periphrastic Possessive with *se* 247

7.4 Comparing Distribution Patterns 251

8 AN INTEGRATED APPROACH TO THE DESCRIPTION OF LINGUISTIC VARIATION AT THE CAPE

8.1 Identifying Lects in the Data
  8.1.1 Multivariate Analysis of Ten Variables 254
  8.1.2 Examining the Social Dimensions 270
  8.1.3 Linguistic Patterns in the Vernacular Writing Tradition 275
  8.1.4 Standardisation and Diglossia Revisited 278

8.2 Code Choice and Code-Switching in the Corpus
  8.2.1 Code-Switching and Linguistic Interference: 285
    Dutch/Afrikaans-English
  8.2.2 Patterns of Code Choice and Code-Switching: 294
    Dutch - Afrikaans

8.3 Die Triomf van Afrikaans 302

EPILOGUE 304
APPENDIX A: STATISTICA - A Short Description of the Computer Package 314

APPENDIX B: The Sample
1 Individuals and Social Information 315
2 The Linguistic Corpus 320
3 The Data Set 326
4 Cluster Membership 332

APPENDIX C
Publication of Afrikaans Books 1861-1925 (number of books per year, based on Nienaber 1948). 334

BIBLIOGRAPHY 335
List of Figures, Maps and Tables

Maps

6.1 Dialectal present tense paradigms in the northern varieties of Dutch (singular only; Van den Berg 1949: 7)

Figures

1.1 The agentivity continuum (adapted from Romaine 1984: 34)
1.2 Structure, social praxis, agency and system reproduction (adapted from Giddens 1991a: 301)
1.3 Sociolinguistic vs. system-oriented approaches to language and language change.
2.1 Model of language change at the Cape (adapted from Raidt 1983: 7, 28)
2.2 Khoe/Dutch interaction: a request made by a Khoe of a European passersby (Smith & Pheiffer 1993: 33)
2.3 Free population and slaves at the Cape, 1692-1793 (in percentages, from table 2.1).
2.4 Demographic balance in plantation creoles (adapted from Bickerton 1995: 317).
2.5 Emergence of Afrikaans according to Den Besten's 'Convergence Model' (adapted from Den Besten 1989: 226).
3.1 Publication of vernacular texts in newspapers and magazines in the Cape Colony 1826-1899 (number of texts per year, based on Nienaber 1966).
3.2 Integration of Afrikaans in the Stammbaum of the Indo-European language family (Du Toit 1891: 3).
4.1 Oral reading practice in the Cape Colony: the daily communal scripture reading (18th century drawing by Egerdörfer; reproduced in Pells 1938).
4.2 Oral reading practice in the Cape Colony: an itinerant teacher reads items from the Zuid Afrikaan (drawing by C. Bells circa 1840; reproduced in Pells 1938)
4.3 The relationship between spoken and written language (Kytö & Rissanen 1983: 473).
4.4 Temporal distribution of the sample texts 1880-1922 (number of texts per year).
4.5 Approximation of the general distribution of the social classes at the Cape (in percentages; based on the census 1904, excluding the category 'peasants').
4.6 The distribution of the sample across the social classes (in percentages).
5.1 Position of cases 1-4 in the two dimensional space.
5.2 Illustration of the Euclidean distance formula (adapted from Kaufman & Rousseeuw 1990: 12).
5.3 Single linkage (from Kaufman & Rousseeuw 1990: 47)
5.4 Complete linkage (from Kaufman & Rousseeuw 1990: 47)
5.5 Tree diagram for single linkage (Euclidean distance).
5.6 Tree diagram for average linkage (Euclidean distance).
5.7 Tree diagram for Ward's method (Euclidean distance).
5.8 Amalgamation graph (Euclidean distance, Ward's method).
5.9 Shepard's diagram.
5.10 Scatterplot (dimension 1 vs. dimension 2).
5.11 Correlation between (NP/del.) and (P/del.) in New York.
5.12 Scree plot.
5.14 Scatterplot (based on component scores).
6.1 Histogram for 136 cases, summarising the degree to which the verbal system of individuals shows loss of person and number distinctions (all verbs, based on percentages for individuals)
6.2 The development of t-apocope according to phonological environment and grammatical constraint (adapted from Conradie 1979: 127-129).
6.3 Apocope of [t] according to phonological and grammatical constraints (based on percentages, mean scores for 12 cases)
6.4 Apocope of [t] according to phonological and grammatical constraints: Johanna Brümmer (based on percentages).
6.5 Possible lines of the development of the apocope of [t].
6.6 Apocope of [t] according to phonological and grammatical constraints: F.S. Malan (based on percentages).
6.7 Apocope of [t] according to phonological and grammatical constraints: P. van Huyssteen (based on percentages).
6.8 Apocope of [t] according to phonological and grammatical constraints: A. Schabort (based on percentages).
6.9 Apocope of [t] according to phonological and grammatical constraints: C. Leipoldt (based on percentages).
6.10 Apocope of [t] according to phonological and grammatical constraints: B.J. Brümmer (based on percentages).
6.11 Apocope of [t] according to phonological and grammatical constraints: Louis Botha (based on percentages)
6.12 Possible lines of development for the regularisation of the past participle of strong verbs.
6.13 Histogram for 100 cases, summarising the degree of generalisation of the present perfect across cases, main verbs only (number of perfect forms as % of the total number of past
tense forms, excluding past perfect forms).

6.14 Histogram for 60 cases, summarising the loss of neuter gender across cases (based on percentages for individuals).

7.1 Histogram for 65 cases, summarising the use of nominative ons across cases (based on percentages for individuals)

7.2 Histogram for 70 cases, summarising the use of het across cases (based on percentages for individuals)

7.3 Histogram for 85 cases, summarising the use of Nie-2 across cases (based on percentages for individuals)

7.4 Histogram for 74 cases, summarising the generalisation of the om...te infinitive across cases (based on percentages for individuals)

8.1 Tree diagram (Ward's method, Euclidean distances) for the variables gender, verb, inf, pp_t, pp_en, tense, pronoun, omte, nie2 and vir.

8.2 Graph of amalgamation schedule for Ward's method.

8.3 Multidimensional scaling (Guttman Lingoes) of the distance matrix for cluster I-III (gender, verb, inf, pp_t, pp_en, tense, pronoun, omte, nie2 and vir)

8.4 Shepard's diagram.

8.5 Plot of means for the three cluster solution (gender, verb, inf, pp_t, pp_en, tense, pronoun, omte, nie2 and vir, based on the results of k-means clustering).

8.6 The disintegration of an existing standard variety: diffusion. The arrow indicates the direction of the development.

8.7 The emergence of a new standard variety: focusing. The arrow indicates the direction of the development.

8.8 Scree Plot.

8.9 Multidimensional scaling (Guttman Lingoes) of the distances matrix for cluster I-III (verb, inf, pp_t, pp_en).

8.10 Multidimensional scaling (Guttman-Lingoes) of the distance matrix for cluster I and II (ons, dit, my/sy, hulle).

8.11 Distribution of 'New' or 'Afrikaans' forms across the three clusters (averaged across the ten variables, in percentages).

8.12 Distribution of social groups across clusters: age (in percentages, from table 8.9).

8.13 Distribution of social groups across clusters: class (in percentages, from table 8.9).

8.14 Distribution of social groups across clusters: sex (in percentages, from table 8.9).

8.15 Multidimensional scaling (Guttman Lingoes) of the distance matrix for cluster I-III (gender, verb, inf, pp_t, pp_en, tense, pronoun, omte, nie2 and vir or 37 cases and nine vernacular texts, based on Euclidean distances)
8.16 Multidimensional scaling (see figure 8.3). Position of the bidualectal speakers F.S. Malan (A[1], A[2]) and Wynanda Hoogenhout (B[1], B[2]).

8.17 Discontinuity: H and L as (variable) diasystems.

8.18 Continuity: Diglossia as a continuum of overlapping linguistic systems.


8.21 Code choices in Wynanda Hoogenhout's letters to her parents (total= 155 letters) 1905-1921 (in percentages).

8.20 Diffusion of Afrikaans forms across time (C 1880-1922). In percentages.

Tables

2.1 Origins of slaves imported to the Cape 1652-1808, in percentages (from Shell 1994: 41).

2.2 Free population and slaves at the Cape, 1692-1793 (from Worden 1985: 11).

2.3 A quantitative model of decreolisation involving shifts in the percentage of speakers using the different lects (adapted from Rickford 1983: 300)

3.1 Wat and die as relative pronouns in the Transvaal periodical press 1866-889 (adapted from Van Rensburg & Combrink 1984: 115).

4.1 A speech community typology based on the degree of recognition of sociolinguistic norms (from Santa Ana & Parodi 1998: 35)

5.1 Two-dimensional data set.

5.2 Distance or dissimilarity matrix for cases 1-4 (Euclidean distances).

5.3 Percentage of full, contracted, and deleted forms of is (Labov 1969: 730).

5.4 Internal structure of cluster II (pre-adolescent peer groups).

5.5 Euclidean distances between clusters.

5.6 Principal components (eigenvalue > 0.05).

5.7 Component loadings (unrotated).

5.8 Component loadings: varimax rotation.

5.9 Component loadings: quartimax rotation.

5.10 Component scores (based on unrotated components).

6.1 Use of Afrikaans forms according to verb class (C-1880-1922).

6.2 Loss of verbal inflection for main verbs (C-1880-1922).

6.3 Loss of verbal inflection for schrijven, leven, regenen, and beginnen (C-1880-1922).

6.4 Loss of verbal inflection for maken and spreken (C-1880-1922).

6.5 Generalisation of is (C-1880-1922).
6.6 Generalisation of *het/ heeft* (C-1880-1922).
6.7 Apocope of -e(n) for plural verbs (present tense), infinitives and past participles (C-1880-1922).
6.8 Apocope of -e(n) for infinitives (C-1880-1922).
6.9 Regularisation of the past participle, group I (/i/, /e/ > /o/ and /ɛ/, /u/, /ie/ > /ʊ/) (C-1880-1922).
6.10 Regularisation of the past participle, group II (/i/, /i/ > /ɛ/) (C-1880-1922).
6.11 Regularisation in past participle for *nemen, beginnen, schrijven* and *krijgen* (C-1880-1922).
6.12 Present perfect-preterite ratio (C-1880-1922).
6.13 Loss of neuter gender in five high frequency nouns (C-1880-1922).
6.14 Frequency patterns for five monosyllabic adjectives (C-1880-1922).
6.15 The emergence of the new system of adjectival declension (C-1880-1922).
7.1 The pronoun system of continental Dutch.
7.2 The pronoun system of Afrikaans.
7.3 *Het, dit and dat* as referential subjects and objects (C-1880-1922).
7.4 *Hulle* as subject and object pronoun (C-1880-1922).
7.5 Summary: The subject/object pronoun variables (C-1880-1922).
7.6 Attribute possessive pronouns in Dutch and Afrikaans.
7.8 *Mij and zij/sij* as possessive pronouns (C-1880-1922).
7.9 *Ons/Onze* as attributive possessives (C-1880-1922).
7.10 The use of relative pronouns following lexical NP antecedents (C-1880-1922).
7.11 Use of objective *vir* with different lexical verbs preceding direct and indirect objects (C-1880-1922).
7.12 Use of objective *vir* with different lexical verbs preceding direct objects only (C-1880-1922).
8.1 Multiple regression (beta correlation coefficient).
8.2 Cluster means for all ten variables as well as the corpus mean (all in %).
8.3 Variance ($\sum (X-M)^2/N-1$) for the three clusters and the corpus.
8.4 Correlation matrix for the ten variables.
8.5 Principal components analysis, factor loadings for factor 1 and factor 2 (GENDER, VERB, INF, PP_T, PP_EN, TENSE, PRONOUN, OMTE, NIE2 and VIR, unrotated).
8.6 Perfect implicational scale (adapted from Romaine 1982: 171).
8.7 Implicational analysis including the variables VERB, INF, PP_T, PP_EN, for 51 cases.
8.8 Implicational analysis including the variables GENDER, VERB, INF, PP_T AND PP-EN, for 40
cases.

8.9 Social characteristics of the three cluster solution (36 individuals, ten linguistic variables).

8.10 Social characteristics of the three clusters solution for morphosyntactic variables (based on four linguistic variables VERB, INF, PP_T and PP_EN, 51 individuals).

8.11 Social characteristics of the two cluster solution for morpholexical variables (based on four linguistic variables ONS, DIT, HULLE, MY/SY, 20 individuals).

Abbreviations

AAVE     African American Vernacular English
ABN      *Algemeen Beschaaafd Nederlands*
adj.     Adjective
Afrik.   Afrikaans
AP       *Die Afrikaanse Patriot*
APO      African Political Organisation
C-1880-1922 Corpus of Private Papers 1880-1922
CDV      Cape Dutch Vernacular
CJ       *Crimineele Justitiën*
CS       Code Switching
df       Degrees of Freedom
Dt.      Dutch
Engl.    English
ET       Extraterritorial
fem.     Feminine
Gmc      Germanic
GRA      *Genootskap van Regte Afrikaners*
IE       Indo-European
KT       *Die Kaapse Taalargief*
KVA      *Kaapse Vernakular-Afrikaans*
masc.    Masculine
ML       Matrix Language
MoI      Medium of Instruction
NGK      *Nederlands Gereformeerde Kerk*
PCA      Principal Components Analysis
OFS      Orange Free State
ORA      Orange River Afrikaans
p  Probability
pl.  Plural
SAL  South African Library, Cape Town
SLA  Second Language Acquisition
SLI  Standard Language Ideology
sg.  Singular
TMA  Tense, Mood, Aspect System
VOC  Vereenigde Oostindische Compagnie
WAT  Woordenboek van die Afrikaanse Taal
ZAT  Zuidafrikaansche Tijdschrift

Symbols
>  bigger than
<  smaller than
≥  bigger or equal than
≤  smaller or equal than
→  becomes, turns into
>  posterior to
~  in variation with
{...}  set of items

Transcription Conventions  Yiddish examples are cited in the standard YIVO orthography following Weinreich 1992 [1949]. No attempt was made to standardise the orthographic variation between Dutch *ij* and Afrikaans *y*, as well as *s* and *sch*. 
Acknowledgements

I am grateful to those who offered their co-operation and support while I was working on this dissertation. Special thanks are due to the staff at the UCT Jagger Library and the UCT Manuscript Collection, the South African Library, the Cape Archives, and to I.H.T. Balie who made material available to me at the Moravian Mission Museum in Genadendal. I am greatly indebted to Karel Schoeman, who not only helped me to find my way through the manuscript collection at the South African Library, but also shared his knowledge and understanding of South African history with me. As a novelist and a social historian, he combines the best of two worlds. Thanks also to the many native speakers of Dutch and Afrikaans who patiently answered my questions and shared their linguistic intuitions with me. I am especially grateful to Carin Dippenaar who spent many hours explaining the finer semantic nuances of Afrikaans idiomatic expressions to me, and to Mr Allie and Mrs van Gend who told me their memories of growing up in Cape Town in the 1920s and 1930s, and tried to remember the languages they spoke and the languages they heard. Discussions with other academics have been important; in particular I would like to thank Carli Coetzee, Achmat Davids, Nigel Love, Carla Luijks, Raj Mesthrie, Johan Oosthuizen, Helma Pasch, Uwe Pörksen, Paul Roberge, Peter Slomanson and Hester Wafer, who have contributed to the gestation of this study by suggesting references and helping to locate inaccessible material, by sharing their own work and thoughts, and by answering questions or commenting on earlier drafts of this dissertation. Many thanks also to Cecily van Gend who gave the manuscript a careful read. I owe a special debt to Brenda Benecke who not only guided me efficiently and cheerfully through the labyrinth of the university administration, but who also provided invaluable personal support. She contributed much to making Cape Town my home away from home. The University of Cape Town provided financial assistance in the form of a Foreign Students Scholarship, the Lestrade Scholarship and a UCT Research Associateship. I would also like to thank my parents Brigitte and Peter Deumert for their support, both intellectually and financially, and their respect for the decisions I have made.

My thinking on the issues discussed in this work has been influenced greatly by my supervisor Roger Lass, and my greatest expression of gratitude must go to him. In innumerable discussions he provided stimulation and encouragement as well as just and constructive criticism, not to mention laughter and emotional support. He continues to be my greatest teacher.

Last, but not least, I want to thank Carel, whose talent for Chinese cooking never failed to lift my spirits in the evenings, who patiently endured my distractedness, and who read through several drafts of this dissertation with the keen eye of a scientist. The responsibility for any errors in the final product remains, of course, my own.

Cape Town, May 1999
PROLOGUE

The interaction of power, language, and reflections on language, inextricably bound up with one another in human history, largely defines language standardization.


This study concerns two complementary social and historical aspects of language: variation and standardisation. While variation is today recognised as a basic condition of language and language change, standardisation has long been neglected by mainstream linguistics, where it was usually treated 'as a non-linguistic subject, the result of laymen's and educator's meddling with the natural processes of language' (Devitt 1989: 1). Standardisation, unlike variation, is not a general condition of natural languages, but rather a historical (never entirely completed) movement towards linguistic uniformity which languages (or dialects) undergo at different times for different reasons. By selecting certain linguistic variants as part of the new norm and by excluding others, standardisation processes reduce linguistic variation, and as such standardisation can be termed a 'complement to variation' (Devitt 1989: 3, see also Haugen's 1972: 107 definition of standard languages as showing 'minimal variation in form' and 'maximal variation in function'). Both variation and standardisation have important functions for the formation of sociolinguistic groups: while divergence (leading to variation) emphasises separateness within the larger community, convergence (leading to standardisation) promotes unity and a common (i.e. supra-local, supra-social) identity. Thus language, as Galli de' Paratesi (1977: 170, cited in Joseph 1987: 42) has put it, fluctuates in its state of natural instability between 'una forza centripeda (la standardizzazione) e una forza centrifuga (la tendenza alla differenziazione}'.

In their study of the prescriptive tradition Jim and Lesley Milroy (1985a) have argued that standardisation processes are usually accompanied by a 'standard language ideology' (SLI): a metalinguistically articulated and hegemonistic belief that there is a correct way of speaking, leading to an ideological 'intolerance of optional variability in language' (Milroy & Milroy 1985a: 1).

---

1 Early work on standardisation emerged from the Prague school of linguistics (see, for example, Havránek 1964 [1938], Garvin 1964), Kloss' (1952, 1978) work on post-1800 Germanic standard languages, Haugen's research on Norway (see the papers collected in Haugen 1972), and an interest in the macrosociological aspects of standardisation processes has always been part of the field of 'language planning'. More specifically linguistic research dates from the early 1980s, often motivated by a general interest in language and power as well as the history of the prescriptive tradition (see for example Haas 1982, Milroy & Milroy 1985a, Joseph 1987, Devitt 1989, Lippi-Green 1994, Stein & Tieken-Boon van Ostade 1994).
However, the papers edited by Stein & Tieken-Boon van Ostade (1994) show clearly that (at least in the case of English) linguistic standardisation understood as the reduction of variation and the emergence of prestige norms, can take place in the absence of a clearly developed SLI. Schiffman (1998) also argued that although such ideologies have played an important role in many standardisation processes they are not a universal condition of standardisation, and similarly Devitt (1989: 2) has noted that 'the two objects - the ideology of standardization and linguistic standardization must be seen as separate entities, for it is quite possible to have one without the other'. In other words, (relative) linguistic uniformity and the emergence of linguistic norms can result from increasing inter-group contact, as the cumulative outcome of a multitude of individual acts of linguistic accommodation, without being accompanied by an SLI. Such processes have been described by LePage & Tabouret-Keller (1985: 181-182, 187) under the term 'focusing'. Although linguistic 'focusing' is not necessarily accompanied by an explicit SLI, issues of power and authority are not absent from the process; their influence, however, is less explicit and will only be recovered by careful analysis. Standardisation through 'focusing' is still not well understood, since it is noticed only retrospectively and usually lacks the kind of overt statements known from cases of more conscious and deliberate intervention.

To capture the distinction between ad hoc processes of cumulative micro-accommodation and convergence (focusing or standardisation in the wider sense), and cases of deliberate intervention (standardisation in the restricted sense), Schiffman (1998: 366-367) distinguished between standardisation through 'informal consensus' and standardisation through 'formal decision making'. A similar distinction was made by Weinreich (1954: 396) in his discussion of standardisation and levelling:

STANDARDIZATION could easily be used to denote a process of more or less conscious, planned and centralized regulation of language. Many European languages have had standardized varieties for centuries; a number of formerly 'colonial' tongues are undergoing the process only now. Not all levelling is equivalent to standardization. In the standardization process there is a division of functions between regulators and followers, a constitution of more or less clear-cut authorities (academies, ministries of education, Sprachvereine, etc.) and of channels of control (schools, special publications, etc.).

---

2 'The Milroys' concept of an SLI elaborates on observations made by Weinreich (1968: 99-100) and Garvin (1964) who argued that the development of standard languages often leads to the emergence of specific attitudes, such as language loyalty, pride and awareness of the norm, while the more general distinction of 'good' versus 'bad' language can also be observed in societies without standard languages (see § 3.3).

3 Recent examples are the emergence of Modern Standard Arabic, a koiné used throughout the Arabic world (Mitchell 1982, 1986), and Standard Spoken Tamil (Schiffman 1998).

4 With respect to the distinction between 'focusing', 'levelling' and 'koinization', Trudgill (1998: 197) writes as follows: 'the reduction of variants that accompanies focusing [...] takes place through the process of koinization. This comprises the process of levelling [...] and the process of simplification.'
Although a distinction between consciously planned standardisation, on the one hand, and focusing through micro-accommodation on the other hand, is conceptually useful, this is not to say that the two types occur in isolation from one another; it is rather that standardisation (understood as the reduction of variation and the development of linguistic norms) is a highly complex process which can stretch over centuries (as was the case for English and German), and usually combines at any stage both (cumulative) micro-accommodation and deliberate intervention. This is particularly true for the late 19th and early 20th centuries, when increasing intergroup contact (as a result of improved transport technology, communication technology, urbanisation and social mobility) supported tendencies towards linguistic focusing (counteracted, however, to some extent by tendencies towards divergence which reflected social conflict and class demarcational needs). At the same time governments as well as individuals increasingly intervened in the process by publishing grammars and dictionaries, by adopting official languages and implementing school curricula, etc. Since deliberate standardisation and processes of focusing often interact, the term 'standardisation' will be used in this study in its wider sense as referring to a linguistic process towards uniformity and stable norms, a process which can be characterised by differing degrees of human intervention.

That important data for a better understanding of standardisation processes could come from detailed case studies has been argued by Haarman (1988: 15-16), Ferguson (1988: 120) and Devitt (1989: 74-75), and both Ferguson and Devitt have emphasised in particular the need for a variationist perspective on standardisation. Variationist studies of standardisation which proceed from the hypothesis that the process of standardisation is reflected in the reduction of the degree of linguistic variation, exist in the work of Devitt (1989) on the diffusion of Anglo-norms in Scotland and in the work of Lippi-Green (1994) on the consonant system in Nuremberg. Not only is a more detailed knowledge of standardisation processes desirable for its own sake, but Devitt (1989) has emphasised that such knowledge is also important for our understanding of linguistic change since standardisation (whether informal or deliberate) is a force which can both inhibit and accelerate evolutive processes of linguistic change.

The study presented here describes the patterns of linguistic heterogeneity in the context of the early standardisation of Afrikaans. The emergence of standard Afrikaans was rapid and strongly marked by cultural nationalism. In 1876 the first grammar was published by the Genootskap van Regte Afrikaners, a language society with a strong nationalist agenda, and by 1925 Afrikaans was recognised as one of the official languages of the country (alongside Dutch and English). Shaffer (1978: 52) has argued that this rather rapid standardisation process

5 Afrikaans is spoken today by about six million South Africans, i.e. roughly 14% of the population (Orkin 1998). This means that after Zulu and Xhosa, Afrikaans is numerically the third strongest language of the country, and Afrikaans is recognised as one of the eleven official languages under
constitutes an ideal object for the investigation and understanding of language standardisation:

Afrikaans provides an ideal case for studying the latter two linguistic developments associated with modernization: vernacular elevation and standardization. For one thing, both processes are very well documented for Afrikaans. For another, vernacular elevation occurred so rapidly in Afrikaans that it's very speed throws into sharp relief the essential characteristics of the process.

Although the history of standard Afrikaans is well documented, descriptions of the period between 1876 and 1925 have so far focused on either overt statements made by nationalist-cultural leaders and other contemporary observers (see, for example, Nienaber & Nienaber 1941, Nienaber 1960, Scholtz 1964, Du Plessis 1986), or on what I have called the 'vernacular writing tradition', i.e. published and deliberately written dialect-literature (see, for example, Lätti 1978, Uys 1983, Van Rensburg & Combrink 1984, Ueckerman 1987, Raidt 1994 [1992]). Comparatively little attention has, however, been paid to linguistic usage in private documents during this time (an exception is De Klerk 1994), and this study thus focuses on a newly collected historical corpus of private papers (1880-1922). The aim is to describe the patterns of variation (and standardisation) observable in these documents and to explore the private/informal linguistic practices at the Cape during a time when deliberate vernacular standardisation was promoted by sections of the speech community. As an example of an empirical, socially-oriented, variationist and quantitative approach to language, this study is based on the framework outlined in Weinreich, Labov & Herzog (1968), while not ignoring later modifications and criticisms which are occasionally grouped under the label 'post-Labovian' (such as, for example, Milroy 1980, Romaine 1982a, Horvath 1985, Dittmar & Schlobinski 1983). The process of standardisation is investigated both from a strictly linguistic perspective (reduction of variability) as well as from a more sociological perspective, focusing on attitudes towards variability as expressed in the

the constitution. Afrikaans is also widely spoken in Namibia where it is used as a mother tongue as well as a lingua franca (especially in the southern and central regions). In term of 'race' or 'ethnicity', over half of its speakers are categorised as 'Coloured', a label which has been applied since the 19th century to descendants of interethnic unions between the indigenous population, the slaves and the European settlers. Modern Afrikaans has a supra-regional standard variety, as well as highly variable social (including ethnic) and regional varieties, showing non-standard features to varying degrees. The focus on written language is inevitable in a historical study, but this is not meant to deny the possibility that similar processes also occurred in the spoken language. In the literature on standardisation the availability of a writing system is usually seen as a necessary condition since it allows the establishment of linguistic models 'across time and space' (Haugen 1972: 246, also Joseph 1987: 6), and the term 'standard language' has often been reserved for written languages. While writing was certainly of great importance in the development of modern standard languages, the universality of this claim has been questioned by Schiffman (1998) based on his work on the emergence of Standard Spoken Tamil, which shows great linguistic uniformity but lacks a written norm (i.e. Literary Tamil still constitutes the written standard).
contemporary press, and the possible emergence of a standard language ideology.

By providing a critical survey of its more recent history I hope to contribute not only to the historiography of Afrikaans, but also to the growing body of research which has increasingly been classified under the label 'historical sociolinguistics' (see, for example, Jahr 1998), i.e. the diachronic study of language use in 'speech communities, social groups, networks and individuals' (Romaine 1982a: x). In chapter 1 the possibilities and limitations of a sociolinguistic perspective on language history are examined. The discussion focuses on the role of speakers as agents in the process of language change, a topic which has been emphasised by many historical linguists in recent years. To describe carefully the scope and constraints of human agency is particularly important in studies of standardisation which show (as argued above) conscious intervention by speakers or groups of speakers to differing degrees. In discussing the problems and possibilities of a sociolinguistic approach to language history, chapter 1 outlines the general theoretical perspective which underlies the remaining chapters of the dissertation. Chapter 2 presents a critical summary of the discussion about the origin and development of Afrikaans; this includes the tradition of Afrikaans scholarship (in particular the work of J. Du P. Scholtz and Edith H. Raidt), as well as more recent approaches which have emphasised the language contact situation in the early settlement and investigated the possibility of a creole ancestor for Afrikaans. Chapter 3, the final chapter of part I, examines the role of the vernacular writing tradition which emerged at the Cape from the 1850s, and summarises the language-political discussions which took place during the 19th century in the context of growing Afrikaner nationalism. Also discussed is the conventional interpretation of the Cape speech community as diglossic as well as the role of power and ideology in the process of vernacular standardisation.

The second part of this thesis is entitled 'Data collection and methodology' and includes two chapters. Chapter 4 offers a detailed sociolinguistic description of the corpus, and a summary of the linguistic and extralinguistic variables investigated. It also includes a critical discussion of some of the central concepts used in sociolinguistic research (in particular the notion of the speech community, and of how to 'measure' the social or extralinguistic variables). An overview of the multivariate clustering techniques used in the empirical analysis is given in chapter 5. Following earlier work by LePage & Tabouret-Keller (1985) as well as Horvath (1985) it is argued that these techniques are promising tools for quantitative analysis since they allow for the description of linguistic variation in terms of (fuzzy) clusters of idiolects (varieties) which are grouped together on the basis of similarity and/or co-occurrence.

The third part of the thesis presents the quantitative analysis of the corpus. Chapters 6 and 7 are item-based descriptions of the different linguistic variables, focusing on their linguistic constraints and frequency distributions in the corpus. Chapter 8 summarises the results of the multivariate analysis which led to the identification of linguistic groups in the data (i.e. groups
of speakers who speak alike). Membership of linguistic groups is then related to the social characteristics of the group members. In the second part of the chapter the variation in the corpus is described from a code-switching perspective, and the analysis shows that the groups identified on the basis of frequency differences also differ from each other with respect to their code-switching patterns. The study concludes with a brief epilogue discussing the implications of the presented material for sociolinguistic theorising in general, and the history of Afrikaans in particular.

Terminology for different cultural and ethnic groups has always been a problem for studies in South African history. I have used capital initial letters (rather than inverted commas) throughout the text when using ethnic labels. This is meant to indicate that these categories should be understood primarily as the ideological constructs of a society obsessed with questions of race and racial categorisation. A few words also on the use of generic pronouns: I decided to make use of the pronoun 'he/him' in its generic sense to avoid stylistically clumsy constructions such as 'he or she', and 'him or her'. However, in cases where it was semantically and syntactically possible, the plural pronouns 'they/them' were used.
PART I

THEORETICAL AND SOCIOHISTORICAL BACKGROUND
HISTORICAL SOCIOLINGUISTICS: SOME THEORETICAL REFLECTIONS

The nineteenth century was the age of history, and the progress which linguistics achieved in espousing the historical point of view was indeed admirable, but the social sciences have established themselves in the meantime, and linguistics must take its place among them as its nature has determined for it. The moment has come to indicate the setting of linguistic problems in accordance with the social point of view.

Antoine Meillet Linguistique historique et linguistique générale (1906)\(^1\)

1.1 Variation Studies, Speakers and Linguistic Evolution

Historical linguistics is centrally concerned with the description and explanation of language change as well as the reconstruction of past language states. Sociolinguistic methods have entered mainstream historical linguistics since the late 1970s primarily through William Labov's work on the interrelationship between linguistic variation and change.\(^2\) Although Labov developed his account of language change on the basis of empirical research in social dialectology, it is not meant to be case-specific, but proposes a general model of the process and nature of language change, and can thus be said to be of 'weak' explanatory value (Lass 1980: 160-169, Aitchison 1987).\(^3\)

Focusing on what Bloomfield (1933) called 'fluctuations in the frequency of forms', Labov (1966, 1991 [1972]) has shown that 'structured heterogeneity', that is, the existence of linguistic variability constrained by linguistic and social environments, is the normal state for any language.\(^4\) The general mechanism of language change as described by Labov can be summarised

---

\(^1\) Translation by W.P. Lehmann (1981: 16).

\(^2\) See, for example the discussion of Labovian sociolinguistics in Hock (1991), McMahon (1994), Hock & Joseph (1996); for an interpretation of Labov's work as constituting the current standard model of linguistic change, see Aitchison (1995).

\(^3\) The question of the nature of linguistic explanations has been discussed in detail by Lass (1980). He (and others) have argued that ('strong') explanations of the D-N type, which dominate in the natural sciences, are an inadequate model for historical explanations since they cannot do justice to the complexity of the subject matter, and imply a deterministic conception of reality which is at odds with the general unpredictability of historical phenomena (also Giddens 1979: 258). Following Heine (1994) I will use the term 'explanation' in the very general sense of 'making sense out of chaos' or 'relief from puzzlement'.

\(^4\) Labov was not the first to show an interest in the linguistic and social constraints on variation. In the tradition of German philology a concern with linguistic heterogeneity and the social embedding of language had already been expressed in the work of Von Gabelentz; within English linguistics one could name, for example, Whitney, Jespersen, Sturtevant, within French linguistics Gauchat and Meillet and
as follows: At any stage language is characterised by variation between competing variants (X, Y, Z...), of a given linguistic variable, which often coexist within the same speaker. Under certain circumstances a linguistic variant X can become associated with membership in a social group A. At this stage the variable is characterised by a variable rule, i.e., a rule whose application is constrained by both linguistic and social factors. Subsequently the rule is extended within group A to new environments, new word classes or segments. As a result the overall frequency of variant X increases, and the variant is now slowly adopted by other social groups leading to a continuous redefinition of its scope to mark group membership (at this stage it can also be redefined as a marker of a new group). As long as there are no opposing social pressures the variant will spread through the entire speech community (and the entire lexicon or grammar), thus gradually ousting the competing variants Y and Z (Weinreich, Labov & Herzog 1968, Labov 1991 [1972]: 178-181). Applying what is known as the 'uniformitarian principle', which states that the forces governing variation and change are the same today as they were in the past, this model has been useful for the discussion of historical processes of linguistic change (such as the discussion of the apparent reversal of the merger involving ME /aː/, /eː/ and /eː/; see Weinreich, Labov & Herzog 1968: 147-150, Labov 1994: 295-302).

As a result of Labov's work on the interrelationship of linguistic variation and change, the study of variation in historical texts has moved from the periphery to the centre of the discipline, and historical linguists have begun to apply the principles of Labovian variation studies to historical texts, leading to 'startling fresh interpretations from the traditional data' (Toon 1983: xii). A variationist perspective was adopted, for example, by Romaine (1982) in her analysis of the relative marker in Middle Scots, by Toon (1983) in his analysis of Old English phonology, by Kytö (1991, also Rissanen & Kytö 1983 and Kytö 1993) for the description of modal auxiliaries in Early American English, by Tieken-Boon van Ostade (1987) in her analysis of the auxiliary do in 18th century English, by Lippi-Green (1994) in her study of the consonantal system in 16th century Nuremberg, and by Joseph & Wallace (1992) in their study of sociolinguistic variation in ancient Rome. To identify the variation space at a given time, and to describe the linguistic and social embedding of the observed variation is usually seen as the main


Although historical linguistics pre 1970 has been primarily concerned with the reconstruction of homogeneous and idealised grammars of past language states, historical linguists have always noticed and recorded the variation found in the sources. However, little or no theoretical importance was attached to such patterns of variation, and surface variation was usually interpreted as a result of dialect mixture and processes of analogy, whose supposedly unsystematic nature was of little interest (for a discussion see Toon 1987, 1991).
objective of 'historical sociolinguistics'. This was emphasised by Mattheier (1987a: 1434):

[...] die vielfältigen Varietäten und Stilstrukturen einer Zeit herauszuarbeiten und jeweils insgesamt als Möglichkeitsrahmen für die danach eingetretene Entwicklung heranzuziehen. Dann werden sich in den manchmal so linear und zielstrebig verlaufenden Sprachgeschichten häufiger Umwege und Holzpfade aufzeigen lassen, die bisher überhaupt nicht in den Blick geraten sind.

Although the variationist paradigm has so far dominated most work commonly described as 'historical sociolinguistics' (as well as sociolinguistic work in general, see Figueroa 1994: 69), there is a second sociohistorical tradition which has focused on macro-sociolinguistic issues such as language attitudes, language policies and language shift. This tradition of research has been referred to as the 'social history of language', and is carried out primarily by social historians with little or no training in linguistics (see the papers in Burke & Porter 1987/1991, also Richter 1985, Burke 1993). The division of labour between 'variationist historical sociolinguistics' and the 'social history of language' is reminiscent of the distinction between 'sociolinguistics (proper)' and the 'sociology of language', which are often glossed as theoretically rooted in linguistics and sociology respectively (Trudgill 1978, Fishman 1968, also Fasold 1984, 1990). However, Hymes (1984: 41-42) has rightly argued that the extent to which a particular researcher will emphasise the more narrowly linguistic or the more general social aspects, is largely influenced by his academic training ('In brief, Ferguson, Gumperz and Labov continued to be certain kinds of linguists. Bernstein, Fishman, Goffman (and Cicourel, Grimshaw) continued to be certain kinds of sociologists'), and the distinction is thus best interpreted as a result of the history of the field, rather than pointing to a true difference in opinion (in this sense see Cameron 1998: 423).

Historical sociolinguistics, like sociolinguistics in general, is thus more than Labovian-type variation studies, and should be seen as including all studies related to the sociohistorical embedding of language and linguistic variation. Furthermore, both sociolinguistics and historical sociolinguistics can be said to be informed by a theoretical perspective which aims at describing and explaining language maintenance and change as the direct result of speaker actions.6 The emphasis on speakers as agents of language change distinguishes historical sociolinguistics from the system-based (i.e. autonomous, internal) accounts of language change which have been typical for much work in historical linguistics, and which have recently been emphasised by linguists arguing for an evolutionary, Darwinian perspective on language history (see, for example, Ritt 1995/1996, Lass 1997a). Advocacy of an evolutionary, speaker-free approach to

---

6 See, for example, Guy's (1990) discussion of language change; with regard to sociolinguistics in general see Hymes' (1974: 75) discussion of Malinowski's statement that 'language [is] a mode of action' as fundamental to sociolinguistic theorising; for a general discussion of sociolinguistics as speaker-centred see Milroy (1992a).
language and language change goes back to August Schleicher who described language as a natural organism which develops in accordance with its given natural laws, without the possibility of human intervention (Schleicher 1863 reprinted in Koerner 1983; note, however, that Schleicher's use of evolutionary terminology was essentially pre-Darwinian, that is static, see Nerlich 1989: 104). While many 19th century philologists found inspiration in the model of biological evolution, the rise of structuralism in the early 20th century was marked by a strong reaction against the use of biological metaphors, and as a result 'evolution', once a powerful paradigm in historical linguistics, 'has become a 'dirty word' in linguistic theory' (McMahon 1994: 314, also Nerlich 1989). However, the current interdisciplinary interest in Darwinian theory leading to 'selectionist thinking [...] extending its tendrils to every aspect of human life, from medicine to morality' (Hurst 1998) has not bypassed linguistics, and attempts to re-establish an evolutionary perspective are slowly gaining support in mainstream linguists. Whether rooted in evolutionary thinking (i.e. seeing language change as proceeding in analogy to the Darwinian processes of replication, mutation, variation and selection) or whether more traditionally structuralist, advocates of system-based accounts of language change generally agree that little (if anything) is gained by placing the speaker's activity at the centre of the academic enterprise; this was argued, for example, by Lass:

By saying we don't 'need' speakers I am not of course making the absurd claim that language change proceeds entirely in their absence. Rather that, as I will argue [...] for certain purposes we don't gain anything by invoking them (whatever their role happens to be), and in fact muddy the waters. Another way of putting this is that there are at least two complementary kinds of historical linguistics: 'structural' and 'psychosocial', say, and this book is about the first kind, which still remains privileged. Even if it can't (yet?) be understood (in terms of the second), it forms a coherent descriptive domain (1997a: 377, fn.42, also Lass 1987a: 171).

Language-internal approaches to language history have certainly shown that language change can be described successfully, with 'historical depth and heuristic richness' (Lass 1992: 255), without explicit references to speaker activity, and (whatever one's theoretical position) as a historical linguist one still writes a history of the Dutch or German language rather than its speakers (Lass 1987a: 154). Before turning in more detail to the question of speakers and speaker action which (as mentioned above) is central to a sociolinguistic perspective, a brief excursus on the rise of evolutionary theories in historical linguistics (already gaining recognition in

---

7 Schleicher took the analogy between linguistics and the natural sciences as being one of identity, a position criticised vehemently by Whitney, who maintained that the relationship could only be one of similarity, but never of identity (Koerner 1993). Subsequent generations of (historical) linguists have generally accepted Whitney's critique and have used concepts borrowed from the natural sciences as metaphors only. This is explicit in Lass (1997a: 293 fn.): 'Here and elsewhere it is important to note that by comparing X and Y as terms of an analogy I am not saying X = Y'.
introductory textbooks, see McMahon 1994) is appropriate since selectionist thinking is relevant to both system-oriented and speaker-oriented approaches towards language change.

Recent evolutionary discussions of language history are indebted to Dawkin's (1989: 189-201) concept of 'memes' which as mental equivalents of genes, are subject to the processes of replication, mutation, variation and selection. Ritt (1995) has suggested that individual phonemes can be conceived of as 'memes', that is, replicators in an evolutionary system. Thus, successful (error-free) replication leads to maintenance and stability, copying errors (mutations) introduce variation, new variants then become the object of selection, and language change can thus be described as the result of cumulative selective retention. Selectionist explanations of the Darwinian type are seductive in their adherence to logical coherence, and most importantly allow for a non-teleological conception of change as variation itself is neutral (being merely the non-functional result of copying errors). While Ritt (1996: 10) conceived of memes as active agents which 'are placing copies of themselves into [a] newly developing brain' (thus verging on a new form of cultural vitalism), Schendel's (1996a) sympathetic critique of Ritt has aimed at reconciling evolutionary and social approaches to language, by arguing that the selection process is governed by social and other extralinguistic factors. In other words, what biologists call 'selective agents' (i.e. environmental factors which favour the selection of a particular variant, see Patterson 1978: 79-83) are often found in the realm of the social (such as the social identity function of a certain variant). Selection pressures furthermore include issues of the speakers' perception and interpretation of the linguistic data, as well as more mechanical aspects such as ease of pronunciation.

A socially oriented interpretation of selection also underlies Keller's (1994) recent account of language change (see § 1.2.3). Like Schendel, Keller argues (1994: 199-201) that while biological genes are active replicators, linguistic memes do not reproduce themselves, but are reproduced by speakers, and the 'goodness' of a meme (i.e. its likelihood of survival via selection) is at least partially a function of its adequacy in helping an individual to achieve its social objectives.9

---

8 The trouble with memes is that unlike genes they have no direct physical substratum, which makes it difficult to be specific about what exactly is being replicated. Lass (1996, also Nerlich 1989: 108) has thus argued that if the concept of memes is to be constructive for language change, we will need to specify which parts of the linguistic hierarchy (phoneme, morpheme, word, phrase, construction, text) constitute the replicator and thus the unit of selection.

9 Although Keller's account of linguistic evolution grants the speaker a central role in the selection process, Keller acknowledges that the mere transmission of memes is not controlled by the speaker. That is, memes are 'contagious': one cannot protect oneself against their intrusion, and one thus 'knows' a great many words, pronunciation types etc. one never intended to learn. However, 'die Eindringlinge zwingen mich nicht, zu ihrer Vermehrung beizutragen. Für den Unterschied zwischen aktivem und passivem Besitz gibt es in der Natur keine Entsprechung' (1994: 200).
Ein Sprachmem bedient sich nicht des Menschen zum Zwecke seines Reproduktionserfolgs (wie es die Gene tun), sondern Menschen bedienen sich der Sprachmeme zum Zwecke des Kommunikationserfolgs; oder allgemeiner gesagt: des sozialen Erfolgs. (1994: 200-201)

Thus, system-internal accounts of language maintenance and change, although intellectually satisfying in their simplicity and logical soundness, cannot be the whole story. Speakers, their cognitive make-up (determining their perception and interpretation of the linguistic data) and the social context in which they are situated, are vital on the level of selection, and our 'quest for knowledge' cannot be satisfied by ignoring speakers and their linguistic behaviour simply because it is difficult to talk about them in a conceptually clear fashion (Antilla 1993: 48). This, however, is not to say that language or language change is necessarily or even primarily something speakers 'do'. The ontology of language cannot be conceptualised narrowly from either an autonomous, psychological, social or semiotic viewpoint, as language is not one or the other, but many things at the same time.¹⁰ That an ontology of language needs to be tentative has been argued eloquently (and aggressively) by Harris (1997: 241, 256):

That is to say, because language does *not* in any particular instance consist of some determinate set of facts or events lying open to inspection, it falls to the linguistic theorist to develop perspectives from which a certain vision of linguistic phenomena becomes possible. [...] Linguistics is one subject which cannot be mapped out in advance by direct appeal to the concept 'language', since that very concept is part of the subject matter under investigation.

And similarly Lass (1997a: 384-385) has argued in the context of historical linguistics and language change, that the question of what language *is* should be approached in a non-essentialist fashion: 'which (kind of) facet is the best to look at given the particular epistemic game one happens to be playing? [...] The point always is fruitfulness, not 'truth'.

It might be possible to get by with 'ontological minimalism' as advised by Lass (1997a: 377, 385), who suggested conceiving of language(s) as population(s) of variants moving through time; however such a 'minimal' ontology is unable to take up the challenge inherent in Andersen's (1989: 14) implicitly speaker-oriented description of the spread of linguistic change:

It is through innumerable individual acts of innovation - of acceptance, adoption, and acquisition - that any new entity gains currency and enters into competition with traditional entities in the usage of a linguistic community.

¹⁰ The question of what language *is* and from which perspective it should, therefore, be approached, falls in the category of 'antinomies' as discussed by Kant in his *Kritik der reinen Vernunft* (1956 [1787]), i.e. questions or problems that allow for contradictory answers, and each proposition can be defended by apparently irrefutable arguments. 'Antinomies' are essentially insoluble, and need to be approached through critical response rather than the adoption of one or the other answer as 'true'.
We are currently not in a position to say whether a speaker-oriented perspective towards the description and understanding of language and language change will ever take precedence over the established system-internal approaches (as Lass 1997a: 385 has noted 'the best we can say is that the 'causal' role of the speaker or the speaker's 'actions' in certain major field-defining kinds of change is not proved'), but unless we develop ways to talk about the relationship of speakers, action and structure in a conceptually clear and epistemologically fruitful way, we will never know what a speaker-centred theory might look like. In other words, leaving the known and proven terrain of language-internal explanations might be barking up the wrong tree, but this is a risk inherent in any academic enterprise. If, however, we succeed in connecting language change (in principle) to the related but nevertheless distinct domains of speakers, social life and social interaction, our explanations are bound to be 'stronger' than language-internal explanations, as they will 'relate language to parameters that are located outside language structure' (Heine 1994: 259).

1.2 The Agents of Change
1.2.1 Desperately Seeking: A Theory of Action

Unsatisfied with the 'invisible' and 'helpless' speaker in system-based accounts of language change, an increasing number of historical linguists have emphasised the 'real need to bring speakers back into the diachronic picture' (Joseph 1992: 127), and James Milroy's monograph *Linguistic Variation and Change: On the Historical Sociolinguistics of English* (1992a, also Milroy 1993) has given programmatic expression to this perspective (for the general trend see the reviews by Odlin 1994 and Davis 1990). Milroy argues for a speaker-centred and interactional approach to language change which should aim at analysing linguistic change 'as a product of speaker activity in social context' (1992a: 4). Language change, according to Milroy, is 'negotiated in face-to-face encounters' (1992a: 21), constitutes a 'social agreement on the linguistic norm' (*ibid.*), and is thus the result of a consensual understanding between speakers: 'there must be a consensus on the implementation of a change [...] The change would not take place if the speakers did not in some way agree that it should take place' (1993: 216). An approach which sees language change as a type of social change and describes speakers as active and rational agents, has also been advocated by Itkonen and Antilla. Their work is in general less concerned with the social embedding of language, but rather focuses on issues of perception, interpretation and meaning (Antilla 1993: 46).

---

11 Already Whitney conceived of language change as the result of social interaction between speakers. Language, according to Whitney, is a social institution, which is formed, developed and altered 'by the will and consent of men' (1867: 49; see Nerlich 1989 for a general discussion of 19th century speaker-centred linguistic theorising).
Both Antilla and Itkonen proceed from the position that historical linguistics (like history itself) is a hermeneutic science, and is as such centrally concerned with the explanation of (rational) human action (Antilla 1993, Itkonen 1982, 1984). According to Itkonen (1982, 1984) (short-term) language change is brought about as the direct result of goal-oriented (i.e. intentional), rational (although not necessarily conscious) action. General tendencies of linguistic change can be explained by the fact that 'if people act rationally they will act similarly in similar situations' (Itkonen 1984: 212). The rationalist model of action, which goes back to Aristotle's Nichomachean Ethics, is illustrated by Von Wright's (1971: 96-98) practical syllogism which models an anonymous individual's action as a goal-belief-action sequence:

X intends to bring about $b$ (GOAL/ INTENTION)
X considers that he cannot bring about $b$ unless he does $a$. (BELIEF)
Therefore, X sets himself to do $a$. (ACTION)

To 'explain' an action is to uncover the intention(s) and belief(s) which motivate the action, and actions are said to be rational if the beliefs are adequate to the goal (Itkonen 1984: 204, 211). However, since the notion of adequacy is not further clarified in Itkonen's (or Antilla's) work, the statement remains largely meaningless. The unspecific nature of the concept of adequacy is troublesome as Itkonen sees 'rationality' as the basic motivation underlying all human action, and 'rationality' thus assumes, as Antilla (1992: 25) has put it, 'the role of causality in the pre-theoretical sense of making things happen' (see Itkonen 1984: 204 on 'rationality' as an explanatory concept).

Von Wright's practical syllogism can be illustrated as follows (from Nyman 1994: 243):

X intends to be heard in a conversational setting involving background noise.
X considers that he cannot be heard unless he raises his voice.
Therefore, X raises his voice.

Here, X can be said to have performed a rational action (taking a common sense notion of 'adequacy' as a guideline), and we are said to have 'explained' his action by stating his intention and belief. Yet, to illustrate the problem of the notion of 'adequacy' consider the following possible modification of stage two (belief):

---

12 Note, however, that even within historiographic research the role of human agency is far from uncontroversial; see, for example, Anderson (1980).

13 The notion of unconscious yet intentional speaker activity also underlies Labov's work, i.e. individuals are said to adopt certain pronunciations to establish themselves as members of a given social group (see, for example, his discussion of Martha's Vineyard: 1991 (1972): 1-42).
X considers that he cannot be heard, unless he grabs his listener by the sleeve, pulls him into a corner of the room, and shouts at the top of his voice in his left ear while gesticulating wildly.

Is in this case the belief adequate to the goal? And more importantly, will X achieve his goal (being heard) or will his conversation partner simply try to escape from the awkward situation? In other words, does this behaviour qualify as an example of rational action? Or is it rather an example of what Pareto (1935 [1916-1919] vol I: 75ff.) has called 'non-logical conduct', that is actions in which the subjective means-end relation does not correspond to the 'real' or 'objective' means-end relation, and which (according to Pareto) form the majority of human actions? Furthermore, how should one deal with cases where an actor has a goal but is unsure in his belief that his action \( a \) will achieve the goal \( b \), but nevertheless performs action \( a \)? Is such an action then half-rational?\(^4\) How can we 'understand' non-rational actions which do not show a 'realistic' or 'adequate' means-end-relation? And finally what is the explanatory status of 'non-rational' (or according to Pareto's terminology 'non-logical') versus 'rational' action in a speaker-centred theory of language change?

Apart from the question of what qualifies as rational behaviour (and there is little doubt that the notion of 'rationality' is inherently fuzzy and best seen as a graded continuum), and whether our actions can be said to be 'typically rational' (Antilla 1992: 25), the assumption that all human action is 'intentional' is also not unproblematic. The criterion of 'intentionality' is used in the philosophy of action to separate action from behaviour or mere happenings, which are generally seen as unintentional and often express a stimulus-reaction relation. However, that the notion of 'intentionality' as a necessary condition for action can be rather problematic was pointed out by Moya (1990: 11), citing a classic example: 'Marrying his mother was something Oedipus did, though it was not intentional'. Consider also the following examples (which are based on the discussion in Moya 1990: 37-60):

(i) I tripped over a branch.
(ii) It was dark. I switched on the light.
(iii) I stepped on someone's toes while trying to secure a place in a lecturehall.
(iv) A dog marks his territory by urinating.
(v) I crossed my legs.

---

\(^4\) Consider the following example familiar to all of those in the teaching profession: I want to help a weak student to improve his writing skills. (GOAL) I do not really think that writing a detailed assessment of his most recent essay will be of much help as he is unlikely to read it. (NEGATIVE BELIEF) I, nevertheless, write an assessment, hoping (against my better knowledge/experience) that my belief is wrong. (ACTION)
While (i) can be identified as an example of a (necessarily unintentional) happening (i.e. I did not mean to trip over the branch lying in the path, it just happened), and (ii) as an example of an intentional action, the situation is less clear with regard to the remaining three examples. I certainly did step on someone's toes in the lecture hall, but as this doing of mine was not intentional, does this mean that by stepping on someone's toes I did not perform an action of some sort? In action theory such instances of 'non-intentional activity' are usually described as examples of action, simply because it is possible to perform the same activity intentionally.

[A]ctions are as a matter of fact intentional under some description; more precisely, a certain piece of behaviour is an action if, and only if, it is intentional (or intentionally performed) under some description' (Moya 1990: 53) [my emphasis]

Only if a certain 'doing' is not intentional under any description, should it be classified as a case of behaviour or mere happening. Accordingly the description of actions generally works under the assumption that 'intentionality is supposed, unless there is a statement to the contrary [...]. So, if someone who does not know Oedipus' story is told that Oedipus married his mother, he is probably induced to think something that is false, namely that Oedipus married his mother intentionally. A non-misleading description of that action should include a qualification, such as 'with no intention' or 'involuntarily' (Moya 1990: 55). The criterion of 'intentionality' thus formulated would, however, also include animal behaviour which, as in example (iv), is certainly intentional under some description, but at the same time instinctive, following a clear stimulus-reaction pattern which is said to be typical for mere behaviour (for an interpretation of animal behaviour as intentional, see Dennett 1997: 323-331, also Giddens 1976: 77). In other words, the conventional contrast between action (=intentional) and behaviour (=stimulus-reaction relation) is highly problematic at closer inspection, and thus best dispensed with (Giddens 1976: 74). Finally, example (v) refers to an action by habit and one might suggest that this covers probably the majority of linguistic actions leading not only to the stability of the system, but also to the establishment of innovations (on the role of ritualisation, including habituation and automatisation, for the development of language, see Haiman 1994). Although most habitual actions are intentional under some description, they are usually not directly motivated (Hamlyn 1990: 136 describes such actions as 'subintentional', i.e. they are non-intentional without being unintentional).

Not only is the notion of intentionality inherently problematic, but our ability to uncover possible intentions and beliefs underlying a given action is also complicated by the fact that people often don't have introspective access to the cognitive processes (intentions and beliefs) which are said to motivate their actions (as shown in the classic experiment by Meier 1931, reported in Nisbett & Wilson 1977: 240-241). However, when questioned about their actions
people nevertheless volunteer reasons for their actions, and in experiments where the stimulus component was controlled, it has been shown that such 'subjective' reasons for action are often nothing but *post hoc* fabrications (for a detailed discussion of the experimental data see Nisbett & Wilson 1977). The possibility of a contrast between 'objective' and 'subjective' beliefs or motives has consequences for what qualifies as an explanation of a given action, i.e. do we understand an action when we understand the motive assigned to the action by the actor himself, or does understanding an action refer to our ability to construct a plausible intention-belief relation which *could* have motivated the action (as suggested by Weber 1956 [1921]: 9)? The second strategy, however, is problematic as the same physical appearance can represent very differently motivated actions (in terms of both intentionality and belief). This was illustrated by Ryle (1968: 480-482) with the following example:

Two boys fairly swiftly contract the eyelids of their right eyes. In the first boy this is only an involuntarily twitch; but the other is winking conspiratorially to an accomplice. [...] Perhaps being new to the art, he winks rather slowly, contortedly and conspicuously. A third boy, to give malicious amusement to his cronies, parodies this clumsy wink. How does he do this? Well, by contracting his right eyelids in the way in which the clumsy winker had winked.

A theory of language which aims to explain language change as the result of meaningful speaker action thus needs to address the 'double hermeneutic' which characterises the practice of social explanations. The double hermeneutic describes the observation that many of the phenomena which the sociologist (or sociolinguist) tries to interpret are already constituted as meaningful by those who produce them. Giddens (1986: 1986: 374) defined the double hermeneutic as follows:

The intersection of two frames of meaning as a logically necessary part of social science, the meaningful social world as constituted by lay actors and the metalanguages invented by the social scientists; there is a constant 'slippage' from one to the other involved in the practice of the social sciences. (Giddens 1986: 374)

Furthermore, to accurately unravel the intended meaning of a linguistic action is difficult since abduction, rather than deduction or induction, is the mode of reasoning which is said to underlie linguistic agency (Andersen 1973). The term abduction goes back to the work of C.S. Peirce, and describes a type of reasoning which starts from an observed result, then invokes a law and finally infers that a specific case is 'true' (while deduction applies a law to a case and then predicts a result, and induction proceeds from observed cases on the basis of which it establishes a law). Antilla (1989: 404) offered the following formalised 'abductive syllogism':
The surprising fact, C, is observed.
But if A were true, C would be a matter of course.
Hence, there is reason to suspect that A is true.

As Lass (1997a: 335) has pointed out, the problem with abductions is that (unlike deductions) they are difficult to model since they involve 'an irreducibly personal element' which is difficult to reconstruct _a posteriori_ if little (or nothing) is known about the abducer.

Neither the linguist's nor the speaker's abductions can be guaranteed to be 'uniform' in any useful way, so that one cannot, on the basis of one's own abductive behaviour, make a sound attribution to others, especially to those belonging to different times and cultures. (Lass 1997a: 336)

Pointing to some of the problems inherent in the concept of rational, intentional actions is not to deny their special privilege in a theory of action. However, the question is to what extent 'rational' action explanations of the type suggested by Iltkonen and Antilla can actually contribute to our understanding of linguistic change. Important types of actions which do not fall under the 'ideal case' of rationality include issues of attention and self-control, of impulse and intuition, as well as subconscious self-deception. And finally there are those actions which Schelling (1978: 14, 17) called 'contingent', i.e. actions that are not characterised by a means-end relation but which depend primarily on other people's actions, and are thus better described as re-actions. That is, by acting people 'are responding to an environment that consists of people responding to their environment, which consists of people responding to an environment of people's responses.'

Implicit in Von Wright's model is also the assumption that intentions and beliefs 'cause' or influence action in a direct way. However, from the 1930s results from attitude research have shown that the beliefs people hold only rarely translate into action (see Brudner & White 1979, Romaine 1995: 317-319), and it has been argued by Blumer that the very notion of a belief as initiating an action presupposes a 'fallacious picture of human action' (1969: 93).

The idea that the tendency to act determines the act presupposes that action is no more than a release of what is already organized. The tendency when activated is held to go over directly into activity, which it guides and shapes. Against this picture I submit that a realistic analysis of the human act reveals an entirely different picture. The human act is not a release of an already organised tendency; it is a construction built up by the actor. Instead of a direct translation of the tendency into the act there is an intervening process which is responsible for the form and direction taken by the developing act [...] this intervening process is constituted by a flow of self interaction in which the individual indicates various things and objects to himself, defines them, judges them, selects from them, pieces together his selections, and thereby organises himself to act.' (1969: 94-95)

In other words, actions are the result of a complex multi-causal process, and individuals
piece together their line of action step-by-step in social contexts. Cases where a hypothetical original goal truly and directly determines the act to the exclusion of situational demands (such as in the case of a junkie craving heroin) are exceptional amongst humans. To accept Blumer's objection (as well as the 'double hermeneutic' discussed above) also means that we cannot invoke idealised individuals in our explanations, but must concentrate on historical actors, whose actions occur in, and are shaped by a specific sociohistorical context, and it is precisely here that an ethnographically oriented sociolinguistics can contribute much to our understanding of the role of human agency in processes of linguistic change (see Lass 1997a: 363). To assume (like Itkonen and Antilla) the existence of a-historical mental and psychological universalities based on the 'common experience of being human' (Antilla 1989: § 23.9.) which allow us to model human actions in terms of universal goal-belief-action sequences, is also questionable when considering, for example, work on the history of mentalities (such as Ginzburg 1980).  

Finally, to describe actors as 'free agents' (Antilla 1992: 23) ignores the importance not only of situational but also of structural constraints (the linguistic system as well as the social structure), i.e. the degree to which individuals are actually free to act varies (see Popper 1962: 90, Lass 1997a: 367-368; the limits of free will are however acknowledged by Itkonen 1984: 211).  

Romaine (1984: 34, also Dittmar 1996: 146) has formulated a hierarchy of linguistic constraints which limit the speakers' potential for agency (understood as the capability for deliberate intervention), and which can serve as a general guideline:

Speakers exercise more power over, and are more active agents in those aspects of the language system where intrinsically meaningful choice exists. The system is virtually open-ended with respect to pragmatics, i.e. the rules and norms of speaking, which are man-made and socially constituted. The speaker is conversely more patient-like in the domain of phonetics and phonology, where the mechanical, physiological constraints of executing or realizing meaningful utterances

---

15 This, however, is not meant to deny the fact that as humans we do share a common 'metaculture' which is based on species-typical mechanisms (such as, for example, the human capacity to acquire language as well as much of our general cognitive make-up), and which is instantiated in individual cultures. This 'metaculture' also includes psychological adaptations such as child-rearing, sexual behaviour and mate selection (including incest taboos), altruistic behaviour etc. (see Tooby & Cosmides 1992). However, when it comes to our understanding of concrete examples of language change in historical societies we are dealing with aspects of human interaction which lie outside this 'metaculture', that is, the social meaning of linguistic variants which influences maintenance or change is part and parcel of historically specific individual cultures. In other words, while certain aspects of our response to variation might be part of a common heritage (such as to align ourselves with powerful groups), the specific choices (i.e. to identify variant A as prestigious) are part of individual cultures which exist in time and space.

16 Antilla (1992: 22, 23) knows of course that speakers change their 'means of expression to constantly changing needs and shifting preferences' as well as that 'outer circumstances' always affect human action, but he doesn't seem to pay much attention to this and continues to advocate the general applicability of Von Wright's practical syllogism for the description of human action.
via a channel in real time are operative.

Figure 1.1 The agentivity continuum (adapted from Romaine 1984: 34)

In sum, while it is important for historical sociolinguists to clarify the notion of speaker agency and its role in processes of language change, it is necessary that this theoretical integration is not limited to actions performed intentionally by rational actors, but also accounts for other types of actions (non-rational actions, habitual actions, contingent actions), as well as for what Andersen (1989) has called 'fortuitous innovations', i.e. spontaneous and purposeless innovations. At the same time the remarks by Blumer remind us, that any discussion of actors needs to pay due attention to the situational as well as the structural constraints in which actions occur.

While I agree with Antilla (1992: 19) that in history (including language history) 'a good description is explanation', I doubt whether the notion of rational, intentional action alone allows us to arrive at 'good' descriptions. Statements such as the following which present caricatures of almost pathologically hyper-rational and intentional speakers are, to my mind, nowhere near a good narrative of what speakers 'do', and are no less reductionist than system-internal approaches.

Social forces show clearly teleology of purpose. Speakers choose to modify their language to assimilate it to other people's norms. The idea of group membership is adopted first, and it then calls out an effect in pronunciation. Similarly, avoidance of homophony is opted for by speakers [...] parts get selected for an efficient result. (Antilla 1993: 57; see also Milroy's (1992a, 1993) exposition of the relationship between speaker interaction and linguistic change, which likewise presents speakers as almost unnaturally active, negotiating and agreeing on processes of language change)

The historical narrative cannot be phrased in universal terms based on psychological generalities (although they certainly have their place), but needs to pay close attention to the ethnographic context in which speakers act, and to relate to the culturally and historically bound 'experience of language' (Harris 1997: 254). Explanations in terms of context-bound speaker
agency will also have to address the question of how individual actions can lead to changes in a collective structure (i.e. language), and how they contribute to the formation of the emergent structural properties of collective phenomena. While some have described collective structures in additive terms simply as the sum of individual actions (see below), others have argued that not the individuals but the collectivity itself is the creator of such emergent structures, a position prominently associated with the work of Emile Durkheim. Since Labov's (1991 [1972]: 261-267) portrayal of language as a collective social 'fact' shows important similarities to Durkheim's social epistemology and ontology, it is expedient to briefly summarise the main tenets of Durkheimian sociology. 17

1.2.2 The Locus of Action: Collectives vs Individuals

In his treatise Rules of Sociological Method (Les règles de la méthode sociologique, first published in 1895, second revised edition in 1901) Durkheim was concerned with delineating the terrain of the newly established discipline of sociology, and identified a range of special phenomena, so-called 'social facts' (faits sociaux), as the subject matter of sociological enquiry. 'Social facts' are ways of acting, thinking, feeling and being which are established through collective practices, and which are 'external' to the individual members of society on whom they are imposed and whose actions they constrain.

When I perform my duties as a brother, a husband or a citizen and carry out the commitments I have entered into, I fulfill obligations which are defined in law and custom and which are external to myself and my actions. Even when they conform to my own sentiment and when I feel their reality within me, that reality does not cease to be objective, for it is not I who have prescribed these duties; I have received them through education. [...] A social fact is any way of acting, whether fixed or not, capable of exerting over the individual an external constraint; or, which is general over the whole of a given society whilst having an existence of its own, independent of its individual manifestations. (1982 [1901]: 50, 59)

Social facts reside in what Durkheim has called the 'collective consciousness' (conscience collective) or 'group mind' (l'âme collective) which is co-extensive with the social collectivity (social group, society), and cannot be reduced to the individual minds of its members. Social facts cannot be created or changed directly through individual action, but emerge, are reproduced and transformed only through the joint activity of the collective. The sociologist can recognise social facts most clearly by the existence of a legal sanction or a moral belief indicating their

17 Durkheimian influence has also been noted for the work of Saussure and Meillet, the latter of whom was closely linked to the 'Durkheimian circle' in Paris (see Lukes 1994: 172). Note however that Koerner (1973: 52ff.) has argued at length against the interpretation of Saussure as having been influenced by Durkheim, an interpretation put forward by Coseriu (1974: 26ff.), Dinneen (1989) and Botha (1991: 4-10).
coercive nature. More 'diffuse' social facts whose obligatory character is less visible, and which Durkheim has described as 'moods' or 'currents', can be identified by their distribution through society which is clearly patterned ('they assume a shape, a tangible form peculiar to them': 1982 [1901]: 54). Durkheim's conception of social facts has important methodological consequences:

(i) to conceive of social facts as external to individuals implies that they cannot be observed by mere introspection, but require empirical analysis (in other words, although we might have an unspecific idea of their existence, our preconceptions and prejudices stand in the way of our 'scientific' understanding; Durkheim 1982 [1901]: 65, 70, 72-74), and

(ii) since 'social facts' form a reality sui generis they can be treated as 'things', i.e. they can be measured, quantified and subjected to statistical analysis just as objects of the natural world (1982 [1901]: 55, 69).

For Durkheim, the determining cause of a social fact is not to be found among the beliefs or actions of individuals, but 'among antecedent social facts' (1982 [1901]: 134), and to explain a social fact (dependent variable) is to correlate it to another social fact (independent variable). Durkheim termed this method of investigation 'the method of concomitant variations' (1982 [1901]: 151).

The mere parallelism in values through which the two phenomena pass, provided that it has been established in an adequate number of sufficiently varied cases, is proof that a relationship exists between them. [...] Constant concomitance is therefore by itself a law' (1982 [1901]: 151).

Durkheim demonstrated his correlational approach to sociological explanation in Suicide (1952 [1897]), a comparative study of European suicide rates, and national suicide rates were shown to correlate, for example, with religion, age and sex. Suicide rates, according to Durkheim's argument, constitute a social fact which exists independently of its individual

---

18 It is important to understand that 'diffuse' social facts are not merely socially patterned, but are always (at least weakly) coercive. Lass' hypothetical example (1997a: 372) of the socially patterned but socially insignificant clothing behaviour of office workers, illustrates this point, i.e. only if there existed a social 'rule' (however diffuse) or obligation for certain groups to wear certain clothes at certain occasions would it be justified to identify the behaviour as a 'social fact'. As Durkheim states (1982 [1901]: 53): 'I may not be conscious of the pressure that they [the social facts] are exerting upon me, but that pressure makes its presence felt immediately I attempt to struggle against them. If an individual tries to pit himself against one of these collective manifestations, the sentiment he is rejecting will be turned against him.'

19 By describing social facts as 'real' Durkheim did, however, not advocate a crude materialism, and he emphasised in the preface to the second edition of the Rules that to treat social facts as 'things' does not imply that they have material referents (1982 [1901]: 35-36). That is, they are 'real' in the sense that π or the structural properties of fractals are 'real' without being 'material'.

---
manifestations, and interacts with psychological/individual factors in 'driving' individuals to commit suicide ('impelling men to self-destruction' 1952 [1897]: 299). The 'true' causes which give rise to national suicide rates are, however, of a social nature, such as the religious affiliations of the society in question (see Aron 1990 [1967]: 44).

That Labov's academic project (especially his New York city study) shows distinct traces of Durkheimian sociology has been noted in discussions of his work (Bickerton 1971: 461, Botha 1991: 11-13, Williams 1992: 70, Figueroa 1994: 76).\(^\text{20}\) Important parallels to Durkheim's thinking are the assumption that language and variability reside not in the individual but in the group, and the insistence on the fact that language as a collective practice can only be investigated empirically, not by introspection.\(^\text{21}\) The normative homogeneity of evaluation which unifies the speech community (see § 4.1.1) is reminiscent of Durkheim's notion of the 'collective consciousness' in that it constrains (or at least guides) the behaviour of individuals, but is clearly not the creation of any one individual speaker. Similarly Labovian variable rules (i.e. probabilistic rules including extralinguistic factors such as group membership) not only summarise the linguistic patterns observed in the speech community, but are simultaneously believed to constrain the linguistic behaviour of individuals (and thus answer almost perfectly to Durkheim's definition of social facts, i.e. they reside in the collectivity but constrain the individual's behaviour; for a discussion of Labov's determinism see Figueroa 1994: 77, 89).\(^\text{22}\) Labov's approach of establishing co-variation between linguistic variants and extralinguistic 'social facts' (class, sex, etc.) parallels the methodology of Durkheim's suicide study (which also constituted the general model for empirical sociology), as does his use of group scores (averages). However, while Labov refrains from giving a clear motivation for his use of group scores, Durkheim sees the use of the statistical average as being directly rooted in his social

\(^{20}\) It seems justified to identify the investigation of variation and change through correlational methods as the core of Labovian sociolinguistics since it was precisely this research which (justified or not) has long been treated as synonymous with sociolinguistics (see Romaine 1982a: 9, Cameron 1990: 82; see also Chamber's 1995 textbook which equates sociolinguistics largely with correlation studies of the Labovian type). However, Labov's work (as well as sociolinguistic research in general) also includes more ethnographic studies describing, for example, the management of identity through language, narratives and discourse (especially Labov's work on AAVE, see also Labov & Fanshel 1977 for a study using discourse analysis).

\(^{21}\) The supra-individual nature of language was also emphasised by linguists such as Whitney, Meillet and, of course, Saussure; see the overview given by Labov himself (1991 [1972]: 185-186, 261ff.), also Shuy (1990: 184-185).

\(^{22}\) The determinism underlying Labovian sociolinguistics is clearly articulated in Chambers' textbook: 'truly idiosyncratic speakers have never emerged from our researches [...] no sample population to date seems to have included one. For the most part, people sound the way you expect them to sound given the facts about their class, sex, age and region' (1995: 100-101).
ontology. That is, the statistical average is a representation of the group mind in its purest form as it cancels out the irregularities of individual behaviour (Durkheim 1982 [1901]: 55, also Porter 1986: 56, 69). In Labovian variation analysis, on the other hand, the use of group scores has generally been presented as a mere issue of practicality (i.e. a small number of group scores are easier to process than a large number of individual scores; see, for example, Hudson 1996: 177-178). However, it had already been shown by the 1950s that correlations based on group scores (so-called ecological correlations) cannot be validly used as substitutes for individual correlations. In other words, to base a study on ecological correlations implies completely different epistemological and ontological assumptions; that is, the object of investigation is the behaviour of the group itself, and not of the individual group members (see Robinson 1950: 351-357, Selvin 1965, Alker 1969).

While many linguists (following Saussure) would agree with the description of language (in the traditional sense of a language, i.e. Dutch, German etc.) as a supra-individual, collective structure, the question is whether one is equally prepared to accept the methodological interferences of a model which assigns ontological priority to collectives. For Durkheim, collectives (represented in quantitative analysis by group scores and often described in anthropomorphic terms as a social 'body' or 'organism') constitute a reality over and above their individual members, and are identified as the actual source of change (as well as social reproduction). Similarly Labov has argued in his New York city study that different social classes (represented in the analysis by a single group score and thus treated as a collectivity) play specific roles in the process of linguistic change; the lower middle class, for example, showing patterns of hypercorrection, is said to propagate linguistic changes from above (see Labov 1991 [1972]: 122ff.). Both Durkheim and Labov have failed to propose a micro-theory based on the actions and behaviour of individuals to back up their collective accounts (i.e. showing how the collective patterns arise as the result of a multitude of individual actions), and both have rejected attempts to account for the phenomena under investigation from an individualistic and psychological point of view (see Durkheim 1982 [1901]: 125ff., as well as Labov's refutation of Paul's individualistic and psychological account of language and language change, 1991 [1972]: 261-262).

The Durkheimian postulate of an independent reality of collectives has been met with much scepticism by later generations of sociologists as the 'group mind' has no material referent and thus escapes direct observation. This position has been articulated clearly in The Structure of Social Action (1937) where Parsons dismissed Durkheim's notion of the group mind as metaphysical, and thus as unsuitable for scientific inquiry.

Since it cannot be observed, it would seem to be a metaphysical entity. And since only observable things are capable of scientific treatment, this metaphysical entity is not a proper object of science. It is a psychic entity, a 'mind'. In so far as minds are observable at all it is obviously only the minds of individuals. The 'group mind', on the other hand, is merely a metaphysical assumption;
its employment is scientifically unsound.' (1937: 155; see also Johnson, Dandeker & Ashworth (1984:156-157) and their description of Durkheim as an 'objective idealist' in the tradition of Plato and Hegel).

The debate between methodological holism and methodological individualism is an old one, but was revived in the 1950s and 1960s by Hayek, Popper and Watkins, all of whom stood up for the defense of methodological individualism (for a detailed bibliography of the discussion see Lukes 1973: 133, fn.11). Since the 1960s sociologists in general have taken an increasingly critical view of methodological holism and its concern with supra-individual structures, and have focused in their theorising more and more on individuals as agents and creators of social structures. The position of methodological individualism is prominently associated with the work of Weber and his project of 'interpretative sociology' (verstehende Soziologie). Although Weber acknowledges that it is often convenient to treat social groups as if they were individual persons who are capable of performing actions (which is reflected in our use of vernacular concepts of collectivity, as in 'the university thinks that...'), he does not assign any reality to such collectives, and regards them as bogus scientifically. A similar position was advocated by Popper in The Open Society and its Enemies where he stated that

all social phenomena and especially the functioning of all social institutions, should always be understood as resulting from the decisions, actions, attitudes, etc. of human individuals [...] we should never be satisfied by an explanation in terms of so-called collectives (1962, vol. II: 98)

Note that a position of methodological individualism does not necessarily combine with 'psychologism' (i.e. insistence on individualistic psychological explanations), and Popper vehemently opposed interpretations of action in terms of (universal) human psychology, insisting on the importance of the social situation (and its constraints) for an adequate interpretation of individual action.

In sociolinguistics, a position of methodological individualism was advocated early by Bickerton (1971) and Bailey (1973, 1980) as well as by the Milroys in their Belfast research (Milroy 1980, also Horvath 1985, Macaulay 1977, Johnstone & Bean 1997, Johnstone 1999). While acknowledging that collective approaches to linguistic variation have produced important results, Milroy (1980: 19) argues that speakers need to be seen as active agents and not as 'cultural dopes' whose behaviour is determined by external constraints such as group or class membership. Seeing individual speakers as active agents also lies at the heart of accommodation theory, a field pioneered by Giles and his associates (for an overview see Giles & Coupland 1991). Criticising Labovian sociolinguistics as taxonomic, descriptive, non-explanatory and deterministic (i.e. not acknowledging the individual as a creative actor), accommodation theorists have focused on the analysis of face-to-face encounters between speakers, and accommodation
theory predicts that depending on the nature of the interaction speakers will change or modify their language either to reduce or to emphasise the differences between themselves and their interlocutor (convergence vs. divergence). Apart from a general interest in speakers as active agents making situated choices on the basis of the linguistic resources available to them, accommodation theory is centrally concerned with the actual situation in which individuals carry out their linguistic actions. This focus on the situation in which the interaction takes places (which is associated in sociology with the work of Goffman and Garfinkel) has been described as methodological interactionism (Knorr-Cetina 1981: 9). Detailed empirical research has given much support to the premises of accommodation theory and has shown speakers as creative (albeit unconscious) actors (see, for example, Bell 1984, Coupland 1984, Großkopf, Barden & Auer 1995).

The position of methodological individualism, although conspicuously commonsensical, is more problematic than often acknowledged. While it is certainly correct to argue that the notion of 'collectives' stands in need of explanation, the same is true for the term 'individual'. If individual is meant to indicate nothing but 'human being' then to see society as an aggregate of individuals is self-evident and trivial (Lukes 1977). If, however, individual is meant to mean 'agent', in the sense of someone contributing to the formation and change of macro-sociological structures, it is necessary to develop a more detailed account of what human agency is and how one gets from individual actions to collective structures which are the explanandum of sociological (or for that matter sociolinguistic) enquiry (see Lukes 1973: 117, Giddens 1979: 94-95, 1986: 220). A simple additive approach was advocated by Collins (1981) who proposed seeing 'macrostructures as nothing but aggregates of micro-structures' (ibid.: 1011, also Boudon 1986); however it is questionable whether collective processes and the emergent properties which characterise collectively situated systems, can, without loss of meaning, be reduced to statements about individuals and their actions (Hund 1982: 274, Lukes 1973: 116).

In sum, the conception of a micro-macro distinction which allows for the analytical separation of the reality of acting individuals from that of larger structures (social institutions, collectives), has been pervasive in sociology and in a reductionist fashion preference has been assigned either to the actions of the individual actors or to the structural properties of larger systems. Since the 1980s, however, the very existence of such a distinction has come under critique and a 'new theoretical movement' has attempted to side-step the 'fake' dichotomy between micro and macro, action and structure by integrating both dimensions into one theoretical framework (for an overview see Alexander 1988, Ritzer 1990). An important contribution to the development of such a dialectical approach is Giddens' theory of structuration (1979, 1986) which
is generally accepted as a substantial contribution to contemporary social theory.\(^{23}\)

1.2.3 Structure, Social Praxis and the Unintended Consequences of Action

To describe the exercise of human agency as being intimately connected to the existence and (re)production of collectively constituted structures is the central concern of structuration theory. In his attempt to rethink the dichotomy between individual action and collective structure, Giddens assigns central status to the notion of *social praxis*, which describes regularities of conduct enacted by individual actors.\(^{24}\) By focusing on the *making* of social life through the enactment of social practices, structuration theory, although centrally concerned with questions of action and agency (which is defined as the general capacity to produce social practices), manages not to lapse into a new subjectivism.

The basic domain of study of the social sciences, according to the theory of structuration, is neither the experience of the individual actor, nor the existence of any form of societal totality, but social practices ordered across time and space. (Giddens 1986: 2)

In the conduct of social praxis, structure and individual are not separated by an ontological gap, but are seen as two sides of the same coin; instead of forming a dualism, they are conceived of as a duality: praxis presupposes structure and *vice versa* (Giddens 1979: 5). The image which comes to mind is that of a spiral in which structure generates praxis, praxis reproduces (or modifies) structure which in turn generates new practices, and so forth.\(^{25}\)

By the duality of structure, I mean the essential recursiveness of social life, as constituted by social practices: structure is both medium and outcome of the reproduction of practices. Structure enters simultaneously into the constitution of the agent and social practices, and 'exists' in the generating moments of this constitution (Giddens 1979: 5)

---

\(^{23}\) An attempt to unify action and structure underlies also Bourdieu's notion of 'habitus' (1984 [1979], see §4.3.2) and Sztopka's 'theory of social becoming' (1991).

\(^{24}\) Giddens developed his theoretical approach through a critical reading of 19th century social theory (Durkheim, Marx, Weber). He also draws heavily on the theoretical and methodological approaches developed by Goffman, Garfinkel and the school of symbolic interactionism, and their focus on agents, social action and the interaction-order. Craib (1992) has thus described structuration theory as a 'theoretical omelette', a synthesis of many different theoretical approaches.

\(^{25}\) Consider in this context Coseriu's (1974: 25) discussion of language: 'Die Sprache erscheint nur im Sprechen der Individuen, und sprechen ist immer eine Sprache sprechen. Das ganze Sein der Sprache bewegt sich notwendigerweise in diesem Zirkel. [...] man darf den Zirkel nicht verlassen, da es sich bei ihm um den Kreislauf der Sprachwirklichkeit selbst handelt und nichts dazu berechtigt einen der beiden Pole als primär zu betrachten.' For Coseriu Saussure's separation of *langue* and *parole* constitutes thus an unacceptable attempt to leave the recursive circle which in itself defines the ontology of language (*ibid.*).
Giddens' definition of structure differs from the way the term is commonly used in the sociological literature, which describes the patterning of social relationships between individuals (i.e. social institutions, see Giddens 1981: 169). In the terminology of structuration theory, the term 'structure' refers to a set of rules and resources which enter into the formation and reproduction of social praxis and social institutions or 'systems' (understood as clusterings of reproduced social practices across time and space). Giddens' analysis of rules and resources can be summarised as follows:

(i) Rules are generalisable procedures which enable individuals to take part in an 'indeterminate range of social circumstances' (1986: 22); rules are trans situational properties of the collectivity and are instantiated through social praxis.
(ii) Resources reflect the power and domination relations in society and as such determine the individual's potential for action. Giddens (1986: 258) distinguishes between allocative resources (such as raw material, land, means of production) and authoritative resources (non-material, such as social status or life chances).  

The concept of resources is important as it allows one to distinguish different types of actors in terms of their capabilities or potential for action, and Mouzelis (1995: 16-18) has argued that the actions of so-called 'mega-actors', i.e. individuals in control of considerable resources, can affect the macro-dimensions of social life directly (also Giddens 1986: 27-28, 1991a: 309). While resources refer to the conditions which influence our potential for action (and thus our participation in the reproduction or change of social practices), rules are said to 'generate' social practices and in turn can be modified only through the emergence of new collective social practices (and not through individual action). Not all rules are, however, equally amenable to change or modification, and Giddens comments that rules relating to the maintenance of social institutions which are deeply embedded in the workings of society (such as marriage, child-rearing, religion and for that matter language) are on the whole more resistant to change than others (Giddens 1986: 171).

Giddens' notion of rules has been criticised for its vague and ambiguous character (see, for example, Thompson 1984: 156) and to date he has failed to provide a clear and consistent account of the concept of rules.  

---

26 Giddens' discussion of authoritative resources is reminiscent of Weber's notion of Herrschaft, i.e. 'die Chance für einen Befehl bestimmten Inhalts bei angebaren Personen Gehorsam zu finden' (1956 [1921]: 28).

27 Giddens has been criticised severely for his superficial account of the degree of freedom that different rules impose on social actors (Thompson 1984, Archer 1991), and Boudon (1982: 7) emphasised
albeit critically (see below) on Wittgenstein's game analogies, i.e. to know the rules of 'game' is to know how to continue it, that is, to know 'how to go on' (see Thompson 1984: 157). In other words, as competent members of a society we know countless rules without necessarily being aware of them or being able to state them explicitly; in this sense they are what Giddens calls 'unacknowledged conditions' of our actions. These rules enable us to 'go on', to participate adequately in social interactions. However, social rules ( unlike the rules of games) are always contested, they are subject to rival interpretations and are continuously transformed through the material and social power implicit in the notion of 'resources'.

Giddens (1976: 118-119), whose theorising shows traces of Saussure, has compared his notion of structures to Saussure's concept of langue, i.e. an abstract and virtual structure which underlies both the (regular, habitual) actions of individuals (i.e. social practices in the durée of everyday life), and the formation and reproduction of systems (i.e. social institutions) stretching across time and space, thus linking the micro and macro levels of social analysis.28

![Figure 1.2 Structure, social praxis, agency and system reproduction (adapted from Giddens 1991a: 301).](image)

---

correctly that 'if one cannot specify the degree of freedom that social agents, taking structural constraints into account, enjoy in any given situation, one is left with profoundly unsatisfactory social theories'.

28 That linking micro and macro would be important for the development of historical linguistics has been argued by Lass (1997a: 363, fn., see also Itkonen 1982). However, regarding the possibility of such an integration Lass is rather more pessimistic than Giddens: 'The most important theoretical integration that could be performed in historical linguistics is somehow establishing (if possible in principle) a clear and intelligible nexus between short-term individual behaviour and long-term linguistic evolution. [...] I suggest that not only is the time not yet ripe for integration, but that the very attempt might be methodologically harmful, since it obscures the crucial distinction between micro- and macro-phenomena, and tries to force the micro into the macro (or vice versa) at the expense of a clear vision of either, with a total neglect of the role of time.' The question raised by Lass is whether it is worth preserving the micro-macro distinction as an analytical tool, and similarly critics of Giddens have argued that his notion of the duality of structure obscures the analytically important distinction between macro-structures and micro-action (Archer 1991, Craib 1992).
Giddens' discussion of action and agency differs from the analytical tradition of the philosophy of action discussed in § 1.2.1, which emphasises the sequential character of action. Giddens considers action as 'a continuous flow of conduct' (1979: 55, 1986: 17-18) which cannot be broken down into separate beliefs and intentions. Accordingly, Giddens has criticised the fact that the nature of social rules has often been discussed using game analogies. Such analogies are highly misleading since they treat rules as 'isolated formulae, [...] related to particular moves [...]. There is not a singular relation between "an activity" and "a rule", as is sometimes suggested or implied by appeal to statements like "the rule governing the Queen's move" in chess. Activities or practices are brought into being in the context of overlapping and connected sets of rules' (Giddens 1979: 65, emphasis in the original). The same holds for the intentions underlying actions:

'Human action occurs as a durée, a continuous flow of conduct [...] Purposive action is not composed of an aggregate of separate intentions, reasons and motives' (Giddens 1986: 3).

Following the critique of conventional action theories by Blumer (see § 1.2.1), Giddens argues that intentions and beliefs rarely have a direct purchase on action but rather supply an overall context which shapes the individual's conduct to differing degrees. Giddens also pays more than passing attention to habitual actions which he sees as vital for the reproduction of social practices and the systems they organise (1979: 7, 1986: 60ff.).

By rejecting Von Wright's sequential model of human action, Giddens manages to sever the strong link between individual intentionality and agency which has been pervasive in the philosophy of action. Action descriptions in structuration theory focus on what an agent 'does' (i.e. the observable output of his agency) rather than his intentions ('Agency refers to doing', Giddens 1986: 10). The (of course structurally limited) potential of the agent 'to act otherwise' is seen as forming the basis of an understanding of action.

Agency refers not to the intentions people have in doing things but to their capability of doing things in the first place [...]. Agency concerns events of which an individual is the perpetrator, in the sense that the individual could, at any phase in a given sequence of conduct have acted differently. Whatever happened would not have happened if that individual had not intervened. [...] Action depends on the capability of the individual to 'make a difference' to a pre-existing state of affairs or course of events. An agent ceases to be such if he or she loses the capability to 'make a difference', that is, to exercise some sort of power.' (Giddens 1986: 9, 14, also 1976: 77)

29 Giddens maintains that all actions are somehow motivated, but this motivation is rarely direct: 'Motives tend to have a direct purchase on action only in relatively unusual circumstances [...]. For the most part motives supply overall plans or programmes [...] within which a range of conducts is enacted. Most of our day-to-day conduct is not directly motivated.' (1986: 6, see also his discussion of the maintenance of 'ontological security' or 'trust' through habitual actions, 1986: 60-64).
To see human action in terms of 'doing' rather than 'intending' also means that it is possible to adopt the 'third-person point of view' characteristic of all science (see Dennett 1997: 356), rather than a subjective or first-person ontology which aims at reconstructing the intentions of real or hypothetical actors from the perspective of the actors themselves. In other words, structuration theory sees the exercise of agency as logically prior to any concern for the actor's subjective intentions or interpretations.

While the reconstruction of an individual's intentions is of little interest to structuration theory, Giddens attaches much importance to the (conscious and unconscious) knowledgeability of actors which provides an overall background for action, and which shapes the agency of individuals (1976: 71-92, 1979: 72, 255, 1981: 162-163, 1986: 9). In his *Notes on a theory of structuration* (1977: 129) Giddens summarises the nature of this knowledge as follows:

> [W]hat an actor 'knows' when he knows how to sustain social encounters with others within a specific community is how to produce 'acceptable' modes of action, to 'understand' both what he himself says, does and what others say and do, and to make judgements about potentially acceptable forms of activity (also Giddens 1986: 281).  

In other words, actors have a (usually tacit) knowledge of the structures (rules and resources) which enter into their actions, a knowledge which allows them to 'go on' in a diversity of social interactions. According to Giddens, reconstructing the historically situated knowledge of actors is part of the process of sociological explanation from the perspective of structuration theory, while reconstructing an individual's intentions and beliefs is not. Although our knowledge of the social world is usually tacit, it can become part of our discursive knowledge since any social praxis can at any time become part of an overt social discourse and thus the direct object of our discursively motivated actions (see Cohen 1989: 55).  

---

30 In his discussion of social knowledge Giddens draws heavily on Garfinkel's work on 'trust' (1963). In the 1960s Garfinkel and his students conducted several experiments which aimed at breaking down these taken-for-granted ways of acting and showed that individuals had little tolerance towards the violation of social practices. One of the transcripts runs as follows: 'S waved his hand cheerily. S: How are you?; E: How am I in regard to what? My health, my finance, my school work, my peace of mind, my...; S (red in the face and suddenly out of control): Look! I was just trying to be polite. Frankly I don't give a damn how you are.' (S= subject, E= experimenter; Garfinkel 1963: 222). Goffman's description of the implicit rules which govern face-to-face interaction in, for example, *The Presentation of Self in Everyday Life* (1959) provides further examples of such implicit social knowledge.

31 Giddens's discussion of agency is usually referred to as the 'stratification model' of action (1986: 5-8). The first level is the 'unconscious' which provides the general motivations for action (see fn.29), the second level comprises the 'practical consciousness', i.e. our tacit, usually pre-verbal knowledge about the workings of society, that is, we 'know' how to interact with other members of our society in adequate ways, but we cannot express this knowledge discursively. 'Practical consciousness' can be investigated through close ethnographic analysis. The third level is the 'discursive consciousness', which refers to our ability to
A final important aspect of structuration theory and Giddens' attempt to dislocate intentionality from agency is the emphasis placed on the unintended consequences of action which are crucial for social reproduction and macro-level change, and which have been largely neglected by the traditional philosophy of action (1979: 59, 1986: 11-12). Thus, when one purchases an ice-cream, the intention is most certainly not to reproduce the currency system, yet the latter is one of the unintended results of the action. Human action is thus bound by the (usually) unacknowledged conditions of action (rules and resources) and the unintended consequences of action (Giddens 1986: 5). Linguistic reproduction (and change) can also be viewed in terms of unintended consequences.

The reproduction of language [...] as condition and result of the production of speech acts or other forms of communication, is not a motivated phenomenon. [...] The speaking of a language, and therefore its reproduction, is not of course unrelated to the wants of its speakers, and is in some part medium of their realisation. Everyone who speaks a language thus has an interest in the reproduction of that language; but the securing of such reproduction is not generally a motivating force among those language speakers. (Giddens 1979: 218)

A model of language change based on the notion of unintended consequences was proposed by Keller (1989, 1994), who suggested to interpret language as a 'phenomenon of the third kind', that is, a phenomenon which emerges as the cumulative result of individual actions without having been the intention of these actions (such as the currency system in the example above). 'Phenomena of the third kind' are located outside the traditional dichotomy of nature and culture: they are distinguished from natural phenomena (such as mountains, the weather) which emerge in the absence of human action, and artefacts (Esperanto, paintings, etc.), which are deliberately created. According to Keller, explanations of 'phenomena of the third kind' should be genetic; they should show how these phenomena could have come into existence.

reflect correctly or incorrectly upon our actions.

The concept of unintended consequences has been used by a variety of social theorists, such as Machiavelli, Vico, Adam Smith, the Scottish school of moral philosophy, Marx, Engels, Pareto, Sorokin, and others. The renaissance of the concept in 20th century social theory began with Merton's classic paper The unanticipated consequences of social action (1936), and it has since formed a central part of theories of social change. There is, however, a crucial difference in the way Merton integrates the notion of unintended consequences into his functionalist framework and the way the concept is used by contemporary action-theorists. For Merton, unintended consequences are not the result of an individual's action, but a product of the functional workings of the system. In other words, one of the (unintended) consequences of my writing this PhD is that supper doesn't get cooked. After a while my husband steps in and does it in addition to his other tasks. A functionalist would interpret this event as the way in which the family system 'reacts' to the situation, adjusting itself by allocating new roles to its members. For action theorists, this interpretation is not acceptable as a system cannot do anything: 'I do something and the action produces a series of intended and unintended consequences, in reaction to which other people choose to act or not to act' (Craib 1992: 36).
Metaphorically such explanations are called 'invisible-hand explanations', since the emergence of the complex structures which characterise 'phenomena of the third kind' appears to have been guided by an 'invisible hand' (Keller 1994: 95-98).

The only pre-requisite for the emergence of an invisible-hand process is that the actions of individuals show similarities which are cumulative in a certain direction, and the first stage in the formulation of an invisible-hand explanation consists thus of identifying the motives and intentions which 'could' have led to similar actions across a group of individuals. Keller suggests expressing them in the form of universal Gricean maxims, such as 'Go from A to B in such a way that the route chosen is the shortest possible' (thus giving rise to the phenomenon of the beaten path), or with regard to language 'Talk in such as way that you are understood', and 'Talk in such a way that you are socially successful' (see Keller 1994: 127-131). An invisible-hand explanation is necessarily a piece of conjectural history, whose adequacy or 'goodness' depends on its plausibility and cogency, and it is not clear if and on the basis of what evidence one can falsify such explanations (Keller 1994: 102, Mühlhäusler 1991: 77).

Ideally an invisible hand explanation consists of three stages, linking the micro-level (which belongs to the area of psychology) to the macro-level (the explanandum) via the invisible hand process which as the cumulative consequence of similar micro-actions is law-like:

| micro | 1. Condition of antecedents: intentions and motives of the individual actors as well as the general conditions under which they perform their actions (explanans). |
|       | 2. General principles or laws: the description of the process by which the macro-structure arises from the multitude of individual actions (the 'invisible-hand process'). |
| macro | 3. The description of the explanandum, i.e. the structure generated by these actions (Keller 1989: 99, 120, 1994: 99, 101). |

A non-linguistic example of an 'invisible hand process' is the cocktail party situation mentioned in § 1.2.1: according to the maxim 'speak so that you will be understood' all guests at the party will continually raise their voices to be heard over the background noise of other conversations. The cumulative (unintended and law-like) consequence of these actions is a continuing increase of the noise level at the party.

Keller's contribution to the explanation of language change is important as he places the structure of linguistic explanations firmly in the context of social theory, and his application of the notion of unintended consequences of intentional action has been greeted with positive interest by many (see Mühlhäusler 1991, Adamsk-Salaciak 1991, Antilla 1992). Criticisms have, however, been levelled against his selection of illustrating examples all of which refer to changes in lexis/semantics, an area where social (in the sense of speaker-oriented) explanations of change have always been accepted. Apart from questions such as whether a sound change can actually
be 'explained' in any depth by an invisible-hand explanation (for a negative argument see Keller 1994: 213-214), Keller's model also contains some conceptual problems. His discussion of agency (the micro-level) is largely based on the intention-belief-action model discussed in §1.2.1, describing human action as intentional and rationally motivated. However, since the notions of 'intentionality' and 'rationality' are highly problematic, further sociolinguistic theorising should follow Giddens' attempt to separate action from intentionality/rationality (see above), and conceptualise of agency primarily in terms 'doing' (i.e. adopt a third-person point of view). Furthermore, although the concept of unintended consequences is important for the explanation of the emergence of complex social institutions (including language), Keller's rather rigid distinction between intended and unintended consequences fails to capture the complexities of social life, and one is tempted to remark that Keller has created his very own prison cell in the Gefängnis der Dichotomien. Giddens (1991a: 309) has emphasised the need to avoid a simplistic juxtaposition of intended versus unintended consequences, and has argued that intentionality (just as everything else connected to human group life) exists in many shadings or degrees.

Put provocatively: 1) social actors quite commonly know more about what they are up to, and its consequences, than social scientists give them credit for; 2) although the division between what is intended and what is unintended seems on the face of things clear-cut, in the light of the subtle nature of the agents' knowability it turns out to have many shadings.

To illustrate this point Giddens considers certain types of suicidal behaviour (especially as manifest amongst adolescents) which are frequently not characterised by the direct 'intent to die', but are better described as 'a gamble with death', 'in which an appeal to others for help and sympathy is as much part of the person's intentions as self-destruction' (1991a: 312). Thus, if the behaviour of family and friends alters as a result of an individual's unsuccessful suicide attempt, it might not be entirely adequate to group this as an unintended consequence, yet few would be so cynical as to classify it as a directly intended consequence.

Not only is the borderline between intended and unintended consequences often fuzzy, it is also frequently questionable as to where one can draw the line in calling something a consequence of a specific action (Giddens 1991a: 312). Consider the following variation on the suicide example: B feels guilty after a suicide attempt of A; as a result B's relationship with C suffers, as well as his professional performance. C, depressed by her deteriorating relationship with B, joins a therapy-group where she meets D. After a while C starts an affair with D, who decides to leave his wife to be with C; in the meanwhile B, now alone, loses his job, turns to drugs and infects himself with HIV when injecting heroin, and so forth and so forth (this has been referred to as the 'accordion effect of action', see Giddens 1986: 10). Should all these subsequent actions be understood as the unintended consequences of A's suicide attempt? Which of these things have been 'done' (intentionally or unintentionally) by A? Merton (1936: 148, fn.) remarked
that the longer the time interval between action and consequences 'the greater the probability [...] that these circumstances have happened by chance', which *mutatis mutandis* means that we cannot understand them in any meaningful way as the result of the unintended consequences of intentional individual actions.

Both Giddens and Keller have been attacked by their critics for failing to provide examples of how a major structural change can be explained as resulting from the *mélange* of the actor's micro-level actions and their intended and unintended consequences.\(^{33}\) Yet, to demand such a study as the ultimate 'proof' shows a misconception about what social theory is, and Giddens in particular has argued that structuration theory is not meant to form the basis of a distinctive research programme, and that it cannot be applied directly to empirical data (1991a, 1991b, re. the limits of the empirical applicability of Keller's model, see Keller 1994: 206-125).

It [structuration theory] is an attempt to work out an overall ontology of social life, offering concepts that will grasp both the rich texture of human action and the diverse properties of social institutions. Some of these concepts should be useful as *sensitizing devices* for research purposes [...]. But I do not think it useful, as some authors have tried to do, to 'apply' structuration theory as a whole in research projects. Moreover, some researchers have such a good intuitive mastery of the complexities of human social activity that they need little help from abstract theoretical formulations - which, while they might be called for on the level of logic, can seem cumbersome when imported unnecessarily into a given research area. (1991a: 310) [my emphasis]

As Craib (1992: 108) has put it, structuration theory tells us 'what kind of things are out there in the world, not what is happening to or between them; it does not give us anything to test or find out'. Nevertheless, ontology and epistemology touch in many ways, and Giddens (1986: 284-286, 1991a) has suggested three principles which should inform empirical research in the social sciences:

(i) **Contextual sensitivity**, i.e. attention to questions of the interaction-order, time and space, trans-historical generalisations should not be assumed *a priori*.

(ii) **Complexity of human intentionality**, i.e. the researcher should aim at penetrating the knowledge of agents discursively, attempting to (re)construct what the agents (bound in time and space) 'know' (both tacitly and discursively), and pay attention to the shifting relations between the intended and the unintended.

(iii) **Subtlety of the social constraint**, i.e. constraints should be seen as moulds for social action, rather than as strict and deterministic rules.

In short, sufficient attention should be given to the ethnographic dimension without losing sight of the structural constraints (rules and resources), and the result should aim at what Geertz

---

\(^{33}\) Note however that at least in sociology a number of studies have made use of Giddens' framework and have shown individuals to be knowledgeable agents reproducing and transforming the social system through their micro-actions (see for example Willis 1977 and Boden 1994, also Giddens 1986: 281ff.).
(1972) has termed 'thick description', i.e. a long narrative about local knowledge, customs, social rules etc. and their contribution to the reproduction of the macro-dimensions of society (a trend rather noticeable in more recent sociolinguistic work, see for example Eckert 1989).

1.3 Directions for Historical Sociolinguistics

Thirty years ago Weinreich, Labov & Herzog (1968) presented their programme for a variationist and empirical approach to language change, and since then variation studies have claimed their rightful place within mainstream historical linguistics. More recently it has been argued that a sociolinguistic approach to language history needs to be more than simply variationist, and should develop a speaker-centred perspective of the explanation of language change. To draw attention to speakers as actors, however, presupposes the need for a clear understanding of the possibilities and limitations of human agency and its role in the production and change of (collective) linguistic structures, including not only intentional-rational actions, but also non-rational, habitual and contingent actions, as well as the unintended consequences of intentional action, and the structural constraints (both linguistic and social) under which speakers operate. Giddens' structuration theory with its focus on the interplay of structure, praxis and agency provides a useful starting point for talking about speaker agency, and his attention to ethnographic detail is an important corrective for more traditional, philosophical models of action which have focused on the sequential description of action as carried out by idealised, a-historical human actors.

Advocating a variationist and speaker-centred approach to historical linguistics is not to say that this is a better or superior way of analysing the linguistic past. There are many cases where talking about speakers and agency gets us nowhere (as rightly argued in Lass 1997a), and where variation analysis is impossible since we can neither collect adequate corpora (of written data) nor do we know enough about the sociohistorical context necessary to approach the question of (situated) speaker agency with sufficient depth. In such cases a speaker-centred and variationist model can only function as 'a methodological guideline and heuristic aid' (Lass 1976: 225). The limitations of historical sociolinguistics were emphasised by Labov (1972: 100) who was rather dismissive of the possibility of historical studies using a variationist perspective, since historians necessarily work with 'bad data', i.e. historical data cannot be collected systematically and one relies on what has survived by historical accident. Labov (and following him Milroy 1992a, 1993) has thus advocated the study of the past through the present, i.e. contemporary sociolinguistic studies based on adequate spoken language data are seen as essential for the formulation of a model of language change which can then be projected backwards (using the uniformitarian principle). However, the recent compilation and publication of the Helsinki corpus for English (as well as the research based on it) has shown that historical sociolinguistics although laborious is no quixotism (at least not for the time post 1600 when the diversification
and increase of written material began). However, since historical data are always and necessarily fragmentary, the responsible historical linguist has to accept that often one will not be able to reconstruct the social embedding of language and variation with any degree of precision, and the aim can thus only be 'to get as close to the historical truth as the data permit [...] "I don't know" is sometimes the only respectable position to take.' (Thomason 1993: 487, 490).

Figure 1.3 Sociolinguistic vs. system-oriented approaches to language and language change.

Not only the availability of data, but also the nature of the change under investigation determines our conceptual apparatus. Certain relatively simple changes which can be clearly formulated as propositions of the type 'at time \( t \), the inflectional system had features \( A, B, C, \ldots \), and at \( t_2 \) these had become \( A^*, B^*, C^* \ldots \)' might be explicable by investigating the speaker's perceptions, agency and the social embedding of the change (provided we have access to the necessary low-level data). It is, however, questionable whether complex long-term changes such as IE \( p \) to Germanic \( f \), or examples of what Sapir (1921) has termed 'drift', which constitute continuous and gradual processes stretching over several centuries, are even in principle amenable to speaker-centred explanations.  

Two areas of language change are often identified as being of particular interest for historical sociolinguistics and speaker-centred explanations: language contact situations in which speakers are conceptualised as active agents in a complex and untutored language learning process, and linguistic standardisation which is characterised by speaker intervention to differing

---

34 On the limits of action-explanations see Itkonen (1982), Keller (1994: 213-215), Lass (1987a, 1997a: 367); for a speaker-centred approach to drift see Anderson (1990), who describes drift as the gradual modification of variable rules employed by speakers; for the related problem of speakers and long-term diachronic processes of grammaticalisation, see Fischer (1997).
degrees. The study of standardisation processes allows linguists to focus on an aspect of linguistic change which has been largely excluded from mainstream historical linguistics, namely the conscious manipulation of linguistic resources by individual actors (language planning in the widest sense of the term). That the study of deliberate language change is an important subject-matter of historical sociolinguistics was argued by Mattheier (1987a: 1447), and the challenge for historical linguists is to establish the limits under which such conscious manipulation operates (on language planning and language change see Rubin 1977, Ferguson 1983, Bochman 1987, Jahr 1989, Willemyns 1989). Particularly from about 1800 (i.e. with the emergence of language centred nationalism, see § 3.2) deliberate linguistic change propagated by individuals (such as, for example, by Ivar Aasen in Norway) or group action (language societies, language academies) interacted more or less successfully with the ongoing process of evolutive language change. Although such planned innovations did not necessarily affect the majority of the speech community, the emergence of social groups whose casual code became a variety close to the new standard language cannot be dismissed a priori as theoretically uninteresting or even irrelevant, and an improved school system, better access to print media and the development of new media (radio, TV) in the 20th century, supported the general spread of standard norms which have begun to affect traditional dialect speech (see, for example, some of the changes in Dutch dialects discussed in De Rooij 1988).

Another important question which has so far received little attention is the possible interplay between language change and social change. There seems to be a general belief that social change influences linguistic change 'somehow', but little has been done to further our understanding of how specific social changes have affected (or not affected) the linguistic system (Mattheier 1987a: 1433). Do, for example, events such as wars and other major crises which generally lead to an acceleration of processes of social change also leave their imprint on the linguistic system? To what extent did (or can) social changes such as urbanisation, the invention of printing, changes in the system of economic production or the political system affect linguistic practices? Examples of lexical change have been explained successfully with reference to external criteria such as contact or technical innovations, and there is evidence that this can be extended to some syntactic changes, as shown by Giesecke in his analysis of the development of scientific registers in German during the 15th and 16th centuries (Giesecke 1992: 280ff.).

In sum, since speech communities consist of people, we need to find ways to speak about what these people do, how the actions of other people influence or limit their capacity to act, and how their actions contribute to the emergence of macro-structures. At the same time we will have to acknowledge that the nature of our explanations depends on the availability of certain types of data, as well as the phenomena under investigation; to quote the anthropologist Gellner (1968: 268):

History is about chaps. It does not follow that its explanations are always in terms of chaps. Societies are what people do, but social scientists are not biographers en grande série.'
AFRIKAANS HISTORICAL LINGUISTICS: MAIN POSITIONS AND DATA BASIS

Every stage of language is a transition stage, every one just as ordinary as any other [...] and thus [one can see] plainly the fluid transitions of its spatial and temporal differences.

Hugo Schuchardt On Sound Laws: Against the Neogrammarians (1885)

2.1 The Emergence of Afrikaans: Prototypes and Variability

The conventional chronology for the emergence of Afrikaans has been established by what Den Besten (1987b: 67) has called the 'South African Philological School', the leading scholars of which are J. du P. Scholtz and Edith H. Raidt. Unlike earlier scholarship which was preoccupied with speculations about the origin of Afrikaans and based almost exclusively on fragmentary sociohistorical evidence, Scholtz and his pupil Raidt pioneered an empirical approach to the linguistic history of Afrikaans, focusing on the diachronic description of individual aspects of grammar and phonology, while external (i.e. sociohistorical evidence) was only of secondary interest (Scholtz 1980: 34).¹

Linguistic data from before 1700 are sparse. For the early years of the colony there are a number of journals kept during official expeditions as well as the official VOC correspondence. Although these texts represent by and large the norms and variation patterns of continental Dutch, occasional examples of colonial usage (especially lexical items) found their way into these documents (see Kloek 1950: 308-330, Raidt 1991: 119-124, 1983: 30-31). For the 18th century there is the archival material collected and published by L.C. Van Oordt (1947-48, 1949). The corpus contains over 300 official letters and reports sent by fieldcornets (veldwagmeesters) to the local magistrates (landdrosten en heemraaden). Most of the texts stem from the late 18th century, and unfortunately only 12 letters are available for the time before 1770. The corpus is stylistically homogeneous and biographical information is available for the writers, most of whom were respected and relatively well-to-do farmers (for a more detailed description of the corpus see Raidt 1991: 131-135). During the 1960s Van Oordt collected further material (including bills, affidavits, letters, court proceedings etc.) covering the first half of the 18th century. Unlike the data collected in the Kaapse Taalgif the unpublished corpus is stylistically heterogeneous and little or nothing is known about the writers (for a critical assessment of the corpus see Raidt 1983: 36, 1991: 135-136). Additional archival material for the early 18th century exists in the documents written by French immigrants (Franken 1953, Pheiffer 1980).

¹ Others usually named as working within this paradigm are J. Smuts, J.E. Loubser, Jacques Conradie, Fritz A. Poneis and R.H. Pheiffer (see Roberge 1990: 148, fn. 15, 1994b: 92, fn. 4).
Other important sources are the diaries by Johanna Duminy (1797, see Francken 1938) and Louis Trigardt (1836-38, see Smuts 1968; see Puddu 1996, I: 52-58 for a summary of other diary sources), and from about 1860 deliberately written vernacular representations occurred in the periodical press at the Cape. The dialect literature constitutes an important but rather problematic corpus which will be discussed in detail in § 3.1.

The very nature of the historical record has led to a preoccupation with the linguistic usage of the European settlers, as linguistic evidence for other ethnic varieties is either fragmentary or rather late (i.e. mid- to late 19th century, such as the Cape Muslim religious writings, see 3.1, or the diary kept by the Nama leader Hendrik Witbooi, see Van der Zwaan 1986). Isolated examples of 17th century Khoe Dutch are reported in Willem Ten Rhyn's Schediasma de Promontorio Bonae Spei (1686) and Peter Kolbe's Caput Bonae Spei Hodiernum (1719). Khoe and slave varieties of Dutch are also recorded in transcriptions of evidence given in court cases from the late 17th century. More recently Roberge (1994b: 66-67, 1995, see also Van Rensburg 1989: 136) has argued that some rural non-standard dialects, such as Griqua Afrikaans, possibly retained older forms of non-European varieties (on Griqua Afrikaans, see Rademeyer 1938, Van Rensburg 1984).

The fact that research conducted within the philological paradigm has concentrated on the language of the European colonists is, however, only partially a result of the incomplete record. Although both Scholtz (1963, 1980) and Raidt (1983, 1991) acknowledge that different (social and ethnic) varieties must have existed in the past, the investigation and reconstruction of non-European varieties of Dutch has been of secondary interest as such varieties are believed to have been marginal to the development of Afrikaans. Both Scholtz and Raidt interpret the emergence of Afrikaans as the result of an essentially gradual and linear development of 17th century dialectal Dutch, which took place within a contact situation characterised by a large number of L2 speakers (Germans, French, Khoe and slaves, see § 2.2.1). Language contact led to an acceleration of the internal tendencies towards deflection and regularisation, characteristic of many West Germanic languages, but never interrupted the gradual process of 'normal linguistic change'. Within the model increasing grammatical simplification is dated to the early 18th century, leading to a transition period between 1740 and 1775. A relatively stable (spoken) vernacular, showing most of the relevant features of (modern) Afrikaans is believed to have been in existence after 1775 (see figure 2.1). All defining linguistic features are believed to have been fully diffused (linguistically, socially and geographically) by 1850.² Note that the general chronology suggested by Scholtz and Raidt agrees with Trudgill's recent formulation about the process of dialect formation in extraterritorial settings, which assumes relative stabilisation in

² Raidt's datings are slightly more conservative than those advanced by Scholtz, who generally saw the process of change as entirely completed by 1775.
the third generation.³

The first stage involves the initial contact between adult speakers of different regional and social varieties in the new location, involving certain types of accommodation and rural dialect levelling. In the second generation, we must hypothesize, children, having many different models to aim at, will demonstrate considerable inter-individual and intra-individual variability. [...] Children will thus have considerable freedom to select variants from different dialects and form them into new combinations, as well as to develop new intermediate and other interlectal forms. Only subsequently, in the third generation, will the new dialect appear as a stable, crystallised variety as a result of focusing processes just described. (1998: 198)

While Scholtz and Raidt have emphasised inherited dialect structures and evolutive language change as the central 'motor' in the development of Afrikaans, others (Combrink 1978: 72-77, Van Rensburg 1983: 138-39, Kotzé 1994) have stressed the role played by processes of koineisation, which is seen as including (following Siegel 1985) processes of dialect levelling and simplification.

Both diachronically and synchronically three basic varieties of Afrikaans are usually distinguished: Western Cape Afrikaans, Eastern Cape Afrikaans and Orange River Afrikaans (see Van Rensburg 1983). Orange River Afrikaans (ORA) is spoken (primarily by Coloured communities) from north of Cape Town to the Etosha Pan in Namibia, in the southern Free State and in an enclave near Kokstad (Kwa-Zulu Natal). Historical sources for this variety are scarce, but as mentioned above, older forms are believed to have been preserved in contemporary varieties of ORA. Eastern Cape Afrikaans is a historical variety of Afrikaans which developed along the eastern frontier of the colony, and is believed to have formed the dialectal basis of modern standard Afrikaans.⁴ Western Cape Afrikaans is spoken in Cape Town and environment, the Sandveld, Boland, Overberg and the Little Karoo. It is based on the non-standard Dutch spoken in the extremely heterogeneous society of the early settlement (consisting of the VOC personnel, visiting sailors, colonists, slaves and the indigenous Khoi, see § 2.2.1). Cape Vernacular Afrikaans (KVA, Kaapse Vernakular-Afrikaans) which is spoken in and around Cape Town is a subvariety of Western Cape Afrikaans (see De Villiers & Ponelis 1987: 45 for a

---

³ Definitions of the temporal extension of a generation vary between 22 and 36 years. They are based on the biological reproduction time, i.e. the time from the first to the last child plus the age of the women at the birth of the first child minus the time of pregnancy (Mattheier 1987b: 81).

⁴ It seems, however, that Van Rensburg's account of the sociohistorical conditions which led to the formation of Eastern Cape Afrikaans, is too simplistic. Historians have pointed out that farmers at the frontier were not as isolated as is often assumed and maintained regular links with the Cape (Guelke 1989, Penn 1995: 35). Thus, rather than a development in isolation from metropolitan influences (as suggested by Van Rensburg, du Plessis & Klopper 1989: 71), Eastern Cape Afrikaans probably developed in a complex process balancing relative geographic isolation with the continuing influence of metropolitan forms.
description of the phonological properties of KVA). Following Roberge (1992: 31, 46), the term Cape Dutch Vernacular (CDV) will be used to refer to the historical transition varieties between Dutch and Afrikaans, i.e. most of the defining features of modern Afrikaans (such as loss of verbal inflection and gender, restructuring of the pronominal system, etc.) are already noticeable, but not necessarily categorical. Other terms used in Afrikaans scholarship to refer to this transition variety include: early Afrikaans (vreë Afrikaans), Afrikaans Hollands, Kaaps Hollands and non-standard Dutch (nie-standaard Nederlands; Uys 1983, Griessel 1991, Ponelis 1993).

| 17th century Dutch dialects (regional and social) | → | Cape Dutch Vernacular (transition variety) | → | 'Afrikaans' |
| internal language change, koineisation and interference (as a result of L2 acquisition) | internal language change |

| 1652 | 1700 | 1740 | 1775 |

Figure 2.1 Model of language change at the Cape (adapted from Raidt 1983: 7, 28).

Within the research tradition of the South African Philological School, continuing linguistic variation in the historical record after 1800, although described in detail (see Scholtz 1965, also Smuts' 1969 quantitative analysis of Trigardt's diary and other Voortrekker documents), is usually interpreted as an artefact, resulting from the influence of the Dutch book language which is said to have existed in a diglossic relationship with the spoken language. Writers are believed to have consciously avoided the use of variants which belonged to the L(ow) variety, and to have randomly borrowed standard features into their language in formal situations, leading to variation patterns in the written language which did not form part of the spoken vernacular. Thus, in her discussion of the pronoun system in the late 18th century, Raidt (1995: 4.1.2. The question of language naming is a difficult one and often results in imposing boundaries on what in reality constitutes a finely graded continuum. In this study, the term Cape Dutch Vernacular (CDV) is used as a broad cover term, referring to a complex variation continuum which comprises a wide range of varieties and linguistic variants, many of which were not selected into the emerging standard language. The term 'Afrikaans', on the other hand, which surfaced in the mid 19th century (§ 3.1) will be used primarily to refer to the emerging standard variety.

states:

[1]n their written language women used consistently the personal pronoun *ik* (following the Dutch spelling) and *wij* for the plural, instead of *ek* and *ons* as spoken by them.

The paradigmatic assumption of a significant difference between written and spoken language has, however, never been substantiated in any detail by either Scholtz or Raidt. The fact that the Dutch standard language which slowly emerged in the course of the 16th and 17th centuries, was not socially diffused before the late 19th century (see Van der Wal 1992, Ponelis 1993: 123), is generally ignored and overt prestige is assumed for Dutch standard forms. Although it is likely that certain CDV variants were socially marked, not all members of the speech community were necessarily able to replace them in their written language with the 'correct' (or even 'hypercorrect') Dutch form, nor can it be assumed that the evaluation of variants was uniform within the extremely heterogeneous colonial society (for a recent discussion of differences in the evaluation of stigmatised variants within a speech community see Santa Ana & Parodi 1998).

Schooling was rudimentary before 1800⁷, and historical research has shown that literacy levels at the Cape (estimated quantitatively on the basis of signature marks in notarial records) were lower than in the Netherlands and colonial settlement in North America (New Netherlands; see Biewenga 1996, De Wet 1981).⁸ Even amongst those who mastered the technical skills of reading and writing, regular access to the norms of metropolitan Dutch was not necessarily given, since books were an expensive commodity in the 17th and 18th centuries. Biewenga (1996: 116-19) estimated that between 1680 and 1730 only 25% of settler households in Stellenbosch owned books (including the bible and devotional literature), a proportion considerably lower than in the Netherlands (note that the books owned were not necessarily in Dutch, but also often in German and French). There was no printing press or newspapers available in the colony until 1795 (Rossouw 1987: 131). Access to the norms of continental Dutch was provided to some extent by private tutors (so called teacher knechts) who were generally European-born; however, their influence was limited. There were only 130 teacher knechts up to 1795 and their teaching skills were reportedly rather poor (judging from the complaints of settlers; Shell 1994: 19-20).

---

⁷ At the eastern and northern frontier schooling was provided primarily by itinerant tutors and the church which required literacy (the ability to sign one's name and to read the Bible) for acceptance into church membership. In Cape Town and surroundings, however, a number of (primary) schools existed (Ponelis 1993: 44; see § 4.1.2 for the education system in the 19th century).

⁸ Literacy has been estimated at about 50% for the settler community (for the 18th century); Dutch literacy was extremely rare among slaves and Khoi; some slaves were, however, literate in languages other than Dutch (Davids 1991: 73ff.).
Explaining the existence of Dutch forms in the data as a result of efforts to follow an acknowledged prestige norm is not only problematic with regard to the sociohistorical evidence (i.e. low literacy levels, lack of access to the norms of the standard), but as Milroy (1992b: 147) has argued, prestige explanations are generally questionable since they are based on an elitist model of society which claims that the norms favoured by the élite will be favoured by everyone in the society. From what we know about sociolinguistic processes in general, it is likely that new local prestige forms emerged at the Cape from the early 18th century when a colonial identity developed amongst second generation settlers (for the historical background see Katzen 1969: 197).

The assumption that the majority of those who were to able to write had a reliable knowledge of the orthographical and grammatical conventions of standard Dutch (or were in cases of doubt able to consult a dictionary or grammar) is thus difficult to substantiate from the sociohistorical evidence. That under such conditions variation in the written medium can be interpreted as paralleling variation in the spoken language has been argued by Lass (1992, 1997a: 61ff., see also § 4.2 for the situation in the 19th century), and Roberge's (1994a: 156) suggestion that the textual record is at least 'in some sense a written reflex of intermediate forms of spoken language' (emphasis in the original) can be accepted as a working hypothesis. If one takes the linguistic variation observable in the documents as 'real', the continuing variation of virtually all defining features of Afrikaans after 1800 has to be integrated into any attempt to date the emergence of Afrikaans as a cluster of new morphological and syntactic features.

That synchronic variation can indicate a linguistic change in progress is today well-established, and the variationist model of language change (see § 1.1) can be summarised as follows: a change from variant A to B is typically characterised by an intermediate period in which both forms coexist, that is, A → {A,B} → B. Historical linguists have shown that the intermediate stage can stretch over long periods of time, or that variation can be stable without actually leading to language change (Romaine 1982a, Lass 1997a: 184ff.). Since scenarios of (longstanding) variation are not uncommon in the history of language(s), it is useful to distinguish conceptually between the accentuation of a change (i.e. the beginning of the process), the diffusion of a change characterised by patterns of variation, and the completion of a change (a 'new rule'). From the perspective of variation analysis the procedure followed by Scholtz and Raidt in which a change is dated as being completed when the new variants have become 'conspicuous' in their frequency, is disputable as under this criterion a period of slow change or

---

9 However, those typically identified with the South African Philological School are not entirely monolithic in their interpretation of the variation found in the written record. Thus Smuts (1969: 59-60) suggested that the variation in the written language did reflect variation patterns in the spoken language, and Conradie (1979: 79) criticised Scholtz for giving 'te min ruimte vir die moontlikheid dat die variasie van woorde [...] ook variasie in die uitspraak kan verteenwoordig' (see also Franken 1953: 188).
even stable variation might be classified as a completed change (such as, for example, the loss of person and number distinctions in the present tense paradigm, which has been dated to 1740 although variation is still prominent in the first half of the 19th century; see the quantitative analysis by Smuts 1969: 17-20, see also § 6.1.1). Following Roberge (1994b: 29) the chronology established by the South African Philological School is best understood as a 'useful diachronic abstraction' (or a 'macro-story' in Lass' terminology 1997a: 288) describing the development of 'Afrikaans' as a prototype, an idealised 'composite of all defining features' (for a critique of the early datings see also Den Besten 1987b: 68-69). An important task for Afrikaans historical linguists is to clarify the theoretical status of the continuing variation in the sources, that is, whether this variation is just a linguistically meaningless byproduct of a diglossic situation, or whether it can be seen as reflecting regular patterns of (spoken) usage.\textsuperscript{10}

A second major critique of the model of language change as proposed by Scholtz and Raidt, is that comparatively little attention is paid to the contact situation of the 17th and early 18th centuries as far as non-European languages and non-European L2 varieties of Dutch are concerned (see the critique in Den Besten 1987b).\textsuperscript{11} While contact (referring primarily to language and dialect contact within the European settler community) is generally accepted as a factor which accelerated the process of linguistic change (via interference), the well-known and controversial hypothesis of a possible pidgin/creole ancestor of Afrikaans has been described as 'vorwissenschaftliche Spekulation' (Raidt 1983: 45).\textsuperscript{12} However, more recently interest in a possible creole ancestor has been revived inside and outside of South Africa, and historical linguists have increasingly turned their attention to non-European varieties of CDV in order to

\textsuperscript{10} To take the variability in written texts as an indicator for the variation patterns in the spoken language also underlies the recent interest of creolists in reconstructing the early stages of creolisation on the basis of historical documents (see for example Van Rossem & Van der Voort 1993, Baker 1995, Arends & Perl 1995, Muysken 1995, the papers in Arends 1995, Van Rossem & Van der Voort 1996). However, even under conditions of rudimentary literacy skills and limited access to prestige norms, not all observed variation is necessarily meaningful, and it is the job of the linguistic historian to determine through detailed analysis whether the variation is regular (and thus meaningful for the understanding of language change) or sporadic and idiosyncratic (Toon 1991: 119).

\textsuperscript{11} Thus with regard to the early Khoe-Dutch sentences (see § 2.2.1) Raidt (1991: 127) states categorically that 'die staatjies van Khoi-Khoi se pidgin-Nederlands [...] leer ons nie veel omtrent die wording van Afrikaans nie, want ons het hier byna uitsluitend met gebroke taal te make wat nóg met Nederlands nóg met Afrikaans verband hou'.

\textsuperscript{12} Raidt's understanding of 'interference' remains rather opaque (1983: 26, fn. 3). According to Weinreich's (1968: 1) influential definition interference is seen as characterising the speech of bilingual individuals, and not communities. It is not entirely clear how, according to Raidt's hypothesis, these idiosyncratic errors influenced the linguistic patterns of the speech community (see also § 2.2.2).
assess their role in the formation of Afrikaans.\textsuperscript{13}

2.2 The Nature and Results of Language Contact at the Cape
2.2.1 Sociohistorical and Linguistic Evidence 1652-1800

Afrikaans shows the general syntax of continental West-Germanic (excluding Yiddish): underlying SOV with V2 in main clauses and \textit{Satzklammer} (sentence brace) in subordinate clauses, a structure unknown in creole languages (which are usually strictly SVO). Afrikaans has also maintained separable prefix verbs, and its verbal morphology, although extremely simplified, is basically Germanic. That is unlike creole languages Afrikaans does not utilise a TMA system, but has retained the form of the Dutch present perfect (with \textit{het} + ge-verb stem) and the Dutch future tense (with \textit{sal} + infinitive). The lexicon of Afrikaans is predominantly Netherlandic, with some important borrowings from Portuguese, Malay and Khoe. The patterns of derivational morphology are largely in line with Dutch (although there are some interesting differences, see Van Marle 1994, Wafer 1994b), the main innovation being the possibility of reduplication (as in \textit{hulle het vroeg-vroeg aangekomen}; for details see Donaldson 1993: 447ff.). The existence of a cluster of innovative syntactic features (i.e. features without direct Germanic antecedent) coupled with almost total morphological regularisation, the comparatively rapid rate of change (less than 200 years according to the datings of Scholtz and Raidt) as well as the sociohistorical scenario during the early years of the colony have been the main motivations for emphasising the role of language contact in the development of Afrikaans (Thomason & Kaufman 1988: 255). The remainder of this chapter is not meant to settle in any way the continuing debate of the 'creoleness' of Afrikaans, but attempts to summarise the central arguments of this discussion.\textsuperscript{14} In general, the position adopted here agrees with the one articulated by Schneider (1990: 105-106), that is, one should be sceptical of attempts to classify languages whose history has been marked by contact as either extraterritorial dialects, semi-creoles or pidgin/creoles, since these terms indicate more precision 'than a complex reality warrants it'.

Simplistic concepts and questions, such as whether a language variety is a creole or not, with no compromises admitted, should be avoided since, in many cases, the phenomenon is too complex to be grasped with the aid of such simple propositions. [...] The facts and outcomes of language

\textsuperscript{13} The expression 'pidgin/creole' is generally used to refer to those cases in which it is not possible to tell whether the language variety discussed is a pidgin or a creole in the traditional sense, i.e. whether or not it has been nativised (Mufwene 1997: 35).

\textsuperscript{14} The discussion of a possible pidgin/creole origin of Afrikaans had and has significant political and ideological dimensions in the context of South African politics which have been discussed by Roberge (1990, 1992).
contact and mixture are more complicated than can be reflected in simple terminologies and classifications. In linguistic matters, more or less are frequently more appropriate responses than simply yes or no. (Schneider 1990 *ibid.*)

There is historical evidence that a rudimentary mixed trade jargon (involving English and Dutch as lexifier languages) developed from about 1590 when European merchant ships began to anchor regularly at the Cape on their way to East India to restock their supplies, and their crews engaged in barter for cattle and sheep with the indigenous population, the Khoi.\textsuperscript{15} The European visitors made no serious attempts to learn the Khoi language(s) and communication proceeded either through a Dutch/English jargon or through Khoi interpreters who were trained by the Europeans from the early 17th century. An extended jargon possibly developed when the *Haarlem* stranded at the coast in 1647, and the crew stayed for more than a year at the Cape awaiting a ship to return home. Their leader Janssen reported that they had daily contact with Khoi groups on which they depended for cattle and sheep, and who acquired according to Janssen's testimony a minimal Dutch vocabulary (for more details regarding the historical and linguistic evidence of these early encounters see Raven-Hart 1967, Den Besten 1986, 1987a, Nienaber 1994a).

The first European settlers arrived in February 1657, five years after the Dutch joint-stock company *Vereenigde Oostindische Compagnie* (VOC) had established the Cape as a refreshment station for their ships between Amsterdam and Batavia. Continuous immigration led to a steady increase of the settler population in the 17th and early 18th centuries (for demographic details see Poneles 1993: 1ff.). The group of European arrivals comprised not only Dutch settlers, but also many non-matrilectal speakers of Dutch, primarily Germans most of whom came from the Low and Middle German dialect areas (about 35% of the early settlers were of German origin). The almost entirely West Germanic character of the settler community was complemented with the arrival of about 200 French Huguenots who had fled France after the revocation of the Edict of Nantes (1685). An isolated settlement of this linguistically homogenous group was prevented by Governor Simon van der Stel (1679-1707), who pursued (in accordance with the VOC) a policy of linguistic and cultural assimilation to maintain the Dutch character of the colony (see Raudt 1983: 12). Pfeiffer (1980) has shown that the Dutch documents written by the French Huguenots anticipate many features which became later characteristic of CDV, such as

\textsuperscript{15} The indigenous inhabitants of southern Africa, the Khoi ('Hottentots') and the San ('Bushmen'), are usually grouped together under the generic name 'Khoesan'. Their cultures, however, are different: the San are nomadic hunter-gatherers, while the Khoi live in a pastoral economy. Elphick (1972) has shown that the relationships between the two groups were complex as poverty-stricken Khoi were often integrated into San communities and San could 'become' Khoi if they acquired livestock (often through theft; for further details see Elphick & Malherbe 1989, Penn 1995: 43-47). In early colonial times contact was limited to Khoi groups as San groups lived the northern interior.
uninflected infinitives and past participles as well as uninflected attributive adjectives. Yet the documents also show L2 characteristics which did not enter CDV, such as loss of plural marking in the noun and French word order. Morphological reduction is also noticeable in the texts written by Germans, which show loss of neuter gender, loss of verbal concord, and [t]-apocope (Ponelis 1993: 20-21).

Regarding the Dutch dialectal base of Afrikaans, Kloke (1950) has argued that Afrikaans has strong affinities to the southern dialects of Dutch, in particular to the regions around Rotterdam, Delft and Gouda. This regional interpretation was challenged by Scholtz (1963: 232-256) who emphasised the importance of an incipient urban Holland koiné which developed in port cities such as Amsterdam and Utrecht where speakers of diverse Dutch dialects as well as refugees from France and Germany came into contact with each other (also Buccini 1996). Demographic figures for the Dutch element of the settler population show that about 28% of Dutch progenitors (1657-1820) came from Amsterdam, another 50% from North and South Holland (Ponelis 1993: 122), thus indicating the importance of Holland varieties as well as the Amsterdam dialect at the Cape. In sum, Cape colonial Dutch speech within the settler community was characterised by a strong dialectal input (regional as well as social), German and French L2 varieties of Dutch as well as a possible urban koiné dialect, while the use of early standard Dutch was largely limited to the higher VOC officials whose stay at the Cape was temporary. 16

The settler population grew initially through immigration, and between 1657 and 1707 only 7% of the free male population were Cape born. However, in a sample of 426 marriages it was found that during the same period the majority of wives were either of Dutch origin or Cape born (41% Dutch and 39% Cape born, De Wet 1981: 112-116), which suggests that within the family environment matrilectal and colonial speakers of Dutch played an important role (only 10% of wives were of French origin and 9% of German origin). From about the 1730s the settler population was mainly native born, exhibited great stability, was close-knit and at least in the frontier regions not yet strictly socially stratified, although a sizeable European working class (i.e. wage labourers working for freehold farmers, so-called knechts) was slowly emerging (Shell 1994: 19, also Katzen 1969: 232, Guelke 1989: 82).

After initial conflicts and wars between the Khoe and the VOC, integration of the Western Cape Khoe into the socioeconomic structures of the colony was largely achieved by 1700, when more and more Khoe (having lost their livestock and grazing rights) began to work as wage labourers, and Khoe women came to perform domestic duties in the colonists' households (for details see Elphick & Malherbe 1989: 10-18, Shell 1994: 26-34, Penn 1995: 75-76).

16 Although we know little about the social origin of the colonists we can hypothesise that they resembled those recruited by the West India Company, which Buccini 1(996: 40) described as individuals 'on the economic or social fringe'. 
A set of anonymous drawings made around 1700 (see Smith & Pheiffer 1993) gives visual evidence (almost like 'camera shots') of the nature of Dutch/Khoe social (and linguistic) interaction before the smallpox epidemic (1713) which decimated the Khoe population in the Cape, destroyed the last traces of their socioeconomic and political independence, and led to their incorporation into the colonial society under conditions of serfdom (Elphick & Malherbe 1989: 21, Shell 1994: 26ff.).

Figure 2.2 Khoe/Dutch interaction: request made by a Khoe of a European passersby (left side of the drawing; Smith & Pheiffer 1993: 33).

In March 1657 the Cape commandant advised the VOC directors to allow the import of slaves to the Cape as the settlers needed additional labour to cope with the demands of their farms (Shell 1994: 6). Prior to that date there were only few slaves at the Cape, most of whom were employed by the VOC (Boëseken 1977: 5, Shell 1994: 445). Cape slaves came from Africa, India, Indonesia, and Madagascar, and according to Shell (1994: 46) the geographic origins of slaves imported to the Cape between 1652 and 1808 comprises 'a broad palette unparalleled in any other recorded slave population anywhere in the world.'
Table 2.1 Origins of slaves imported to the Cape 1652-1808, in percentages (from Shell 1994: 41). Note that changes in shipping patterns and competition between slave trading companies shaped the import of slaves to the Cape: the first slaves came primarily from Africa, between 1670-1700 most new slaves came from Madagascar, from about 1700 (to 1780) slaves came predominately from India and Indonesia. Import of slaves from Madagascar increased again in the 1750s, and import of African slaves from about 1780 (Shell 1994: 41-46).

<table>
<thead>
<tr>
<th></th>
<th>Africa</th>
<th>India</th>
<th>Indonesia</th>
<th>Madagascar</th>
</tr>
</thead>
<tbody>
<tr>
<td>origin of slaves</td>
<td>26%</td>
<td>26%</td>
<td>23%</td>
<td>25%</td>
</tr>
</tbody>
</table>

The chances for the emergence of a separate slave culture (including the development of a common linguistic code) were slight since no large slave-holders (other than the VOC) existed at the Cape, and the isolated location of many farms made interaction between slave groups difficult (Worden 1984: 32). The slave population was, furthermore, not self-reproducing. Possible social and linguistic stabilisation was thus regularly interrupted by the arrival of new slaves, and only in the 1760s did the percentage of locally born slaves reach the 50% mark (De Wet 1981: 151, Armstrong & Worden 1989: 147-148, Shell 1994: 47). Historians of Cape slavery have emphasised the close contact and intimacy in which slaves and their masters lived together, which supported the assimilation of the slaves into the dominant colonial culture (Shell 1989, 1994: 214-227; note that the situation was similar for Khoi who worked for the colonists, see Sparrman 1787 I: 137). However, the slaves also influenced the domestic culture of their masters, and as a result a rather homogenous colonial culture (predominantly European with some African and Asian traits) developed in the Cape (Elphick & Gilliomee 1989: 225).

Apart from regular and often intense contact through the conditions of bondage (settlers-slaves) or wage labour (settlers-Khoi), sexual encounters were common especially during the early decades of the settlement when the sex ratio among the colonists was strongly biased towards men (Shell 1994: 291). Travellers reported that the majority of children born to slave women were of mixed origin, and European-slave unions were often encouraged by slave owners since children born to slave women were the property of the slave holder, who, thus, stood to gain financially as 'Mulatto slaves' were most valuable (Elphick & Shell 1989: 195, Shell 1994: 56; the VOC, however, tried to discourage such unions). At the eastern and northern frontier of the colony wage-labour relations between colonists and (free) Khoi also shaded frequently into cohabitation (Guelke 1989: 82). 'Mixed marriages' were not uncommon during the 17th and 18th century, and Heese (1984) counted over 1000 marriages between Khoi or ex-slave women and

---

17 The largest concentration of slaves lived in the VOC's slave lodge, and sporadic evidence for a developing slave culture refers to the lodge slaves (Shell 1994: 49). However, from the beginning of the 18th century lodge slaves only formed a very small percentage of the overall number of slaves in the colony.
settlers of European descent between 1660-1799. Miscegenation occurred not only between Khoe/slaves and settlers but also between Khoe and slaves (Giliomee & Elphick 1989: 201). The illegitimate (free) offspring of mixed unions were known as Bastaards, a term which came to indicate mixed parentage as well as a specific social status, i.e. they were generally able to avoid menial labour and often worked as craftsmen or small farmers (Legassick 1989: 370). From this social nucleus a series of communities (such as the Griquas or the Rehoboth Bastaards) emerged in the 18th and 19th centuries along the northern frontier. These communities were largely bilingual (CDV-Khoe) and followed Khoe as well as colonial customs.

Creole Portuguese and Pasar Malay were important lingua francas within the slave community, and knowledge of especially Creole Portuguese has been reported for some settlers and Khoe (see Franken 1953: 41-79, Shell 1994: 59-63). However, since knowledge of these languages was limited outside the slave community, their spread as lingua francas for inter-ethnic communication was prevented (Raidt 1983: 20, Ponelis 1993: 17). The numerical strength of the European element (see table 3.1) furthermore supported the use of a variety of Dutch as the main medium for communication between settlers and slaves.

<table>
<thead>
<tr>
<th>Year</th>
<th>Slaves</th>
<th>Settlers</th>
<th>Settlers % of population</th>
</tr>
</thead>
<tbody>
<tr>
<td>1692</td>
<td>337</td>
<td>799</td>
<td>70%</td>
</tr>
<tr>
<td>1701</td>
<td>891</td>
<td>1,265</td>
<td>59%</td>
</tr>
<tr>
<td>1711</td>
<td>1,771</td>
<td>1,756</td>
<td>50%</td>
</tr>
<tr>
<td>1723</td>
<td>2,922</td>
<td>2,245</td>
<td>43%</td>
</tr>
<tr>
<td>1733</td>
<td>4,709</td>
<td>3,074</td>
<td>40%</td>
</tr>
<tr>
<td>1743</td>
<td>5,361</td>
<td>3,972</td>
<td>43%</td>
</tr>
<tr>
<td>1753</td>
<td>6,045</td>
<td>5,419</td>
<td>47%</td>
</tr>
<tr>
<td>1763</td>
<td>7,211</td>
<td>6,750</td>
<td>48%</td>
</tr>
<tr>
<td>1773</td>
<td>8,902</td>
<td>8,465</td>
<td>49%</td>
</tr>
<tr>
<td>1783</td>
<td>11,950</td>
<td>11,040</td>
<td>48%</td>
</tr>
<tr>
<td>1793</td>
<td>14,747</td>
<td>13,830</td>
<td>48%</td>
</tr>
</tbody>
</table>

Table 2.2 Free population and slaves at the Cape, 1692-1793 (from Worden 1985: 11).
The colonial society at the Cape resembled what Mufwene (1996: 91) has called a *homestead society*, i.e. a society in which the average number of slaves per homestead was small (usually below ten) and interaction between settlers and slaves was frequent and intense. In those territories where prototypical creoles emerged (Caribbean, Gulf of Guinea and Indian Ocean) the *homestead stage* was followed by the *plantation stage*. This second stage was characterised by relatively strict segregation between a small group of European slave holders and large groups of slaves speaking mutually incomprehensible languages which led to the nativisation of the pidgin/creole within the isolated slave society.\(^{18}\) Thus, in Jamaica the proportion of slaves to settlers was roughly 3:1 as early as the 17th century, and 10:1 by the mid 18th century (Mufwene 1996: 92, see also the statistics given by Singler 1995 for Haiti and Martinique). Baker (see, for example, 1993: 137-138) has suggested that the typical scenario under which pidgin/creoles arise is characterised by three demographic events:

Event 1 - when the number of slaves first surpasses the slave-owning population\(^{19}\)
Event 2 - when the number of locally born slaves surpasses the slave-owning population
Event 3 - when the regular supply of slaves comes to an end

Autonomous creoles typically arise after event 2, when the newly arriving slaves target

---

\(^{18}\) See also Bickerton’s (1989) distinction between fort and plantation creoles, and his discussion of degrees of ‘radicalness’ understood as linguistic distance from the lexifier language (Bickerton 1984).

\(^{19}\) Singler (1995: 269, fn.2) and Roberge (1998: 20) define event 1 slightly differently as the moment at which slave and European populations achieve numerical parity.
their language acquisition towards the emerging contact language. The data summarised in table 2.2 and figure 2.1 indicate that event 1 occurred at the beginning of the 18th century (in the year 1717 according to Shell 1994: 4), thus roughly 60 years after initial colonisation. However, even then the slave population was only marginally larger than the settler population. Event 2 never took place: when the percentage of locally born slaves rose to over 70% in the 1830s its numbers did not surpass the settler population (see Shell 1994: 47; note also that the percentage of slave women and children was consistently low at the Cape, see the data in Shell 1994: 48, 446-448). The demographics typically underlying creolisation are summarised in figure 2.4: initial but short-lived numerical dominance of the settler population, quickly (within a few years) superseded by the slave population.

![Graph showing demographic balance in plantation creoles](image)

Figure 2.4 Demographic balance in plantation creoles (adapted from Bickerton 1995: 317).

Bickerton (1981: 4) has suggested that a proportion of about 20% superstrate speakers is the upper limit for creolisation to occur, while higher percentages of superstrate speakers support the gradual acquisition of the superstrate language through the substratum population (see also Thomason & Kaufman 1988: 155-156).\(^{20}\) Similarly Mufwene (1996: 100) emphasised that homestead societies usually don't support the emergence of creoles, but L2 approximations of the dominant language.

---

\(^{20}\) Note that Bickerton (1984: 176) distanced himself later from this rather rigid threshold, describing it as a somewhat 'simplistic approach'. Yet, the basic principle remains, a large percentage of superstrate speakers (possibly between 1/3 and 1/2 of the population) supports processes of L2 acquisition in the substrate community rather than the emergence of a pidgin/creole.
It is very unlikely that anything close to today's creoles was developing on a large scale, even if subsystems close to those of today's creoles may have been shaping up [...] in the speech of some individuals. Rather, approximations of European speech are likely to have been the trend among the non-Europeans living fairly closely with the European colonists. There is no reason why normal people (which the non-European labor generally were) living intimately with speakers of the lexifier would of necessity have developed creoles instead of close approximations of the lexifiers.

And Roberge (1995: 78) states regarding the situation in the Cape:

One cannot assume that stabilisation of the pidgin occurred within the slave community itself, however diverse its linguistic or ethnic composition may have been. Sociolinguistic conditions at the old Cape could only have ensured a greater exposure of slaves to the superstrate language and hence a robust L2 model.

In sum, while the social organisation of plantation societies supported the stabilisation of autonomous pidgins/creoles for slave interaction, the lack of social distance between slave owners and slave(s) characteristic of homestead societies is likely to have led to a dominance of L2 processes, supporting the emergence of indigenised varieties of Dutch at the Cape.

Den Besten (1986: 198), however, who assumes a direct creole ancestor for Afrikaans, has argued that a relatively stable Dutch pidgin developed among the Khoe in the five to six years before the arrival of the slaves. The earliest evidence for Khoe Dutch is recorded in an official journal written by Pieter Meerhoff in 1661 (quoted and discussed in Den Besten 1986: 214-215, 1987a: 15-16). Meerhoff describes in his journal how during a journey into the interior his Khoe interpreter warned him:

(1) Namaqua boeba kros moscoqua
   NAMAOQA OX SKIN VERY-ANGRY
   (The) Namaquas (with their) ox skin shields (are) very angry.
   (The) Namaquas (have) ox skin shields. (They are) very angry. 21

This utterance is certainly not immediately comprehensible to Dutch speakers, but as Meerhoff does not give a translation one can assume that the readers of the journal (i.e. VOC officials in the Cape) were more or less familiar with such utterances (an alternative interpretation is that Meerhoff was trying to impress the VOC officials with his local knowledge and the quotation of the sentence is symbolic rather than informative; the possibility of such an interpretation was suggested by Paul Roberge, p.c. August 1998). From 1671 sentences with SOV

21 The translations are based on Den Besten's detailed discussion (1987a: 15-16). Boeba is a pidgin word for 'ox' attested as early as 1615, kros (< karos) is Khoi for 'skin rug' and moscoqua can be analysed as moeske kwaad.
syntax occur in the small corpus we have of Khoe Dutch. Unlike example (1), however, these utterance are generally mutually intelligible with Dutch.

(2) 1672
't Za lustigh, duitsman een woordt calm, ons V kelum
TZA QUIET DUTCHMAN ONE WORD SAY, WE YOU CUT-THROAT
(Franken 1953: 113)

(3) 1673
Duysman altyt kallom : Icke Hottentots doot makom
DUTCHMAN ALWAYS SAY : I HOTTENTOTS DEAD MAKE
(Ten Rhyne 1686: 140)

The early Khoe Dutch sentences anticipate certain characteristics of Afrikaans, such as the use of the object pronoun ons in subject position, and in the sentences reported by Kolbe the Dutch demonstrative die occurs as definite article as well as a demonstrative pronoun (for a linguistic analysis of the Khoe Dutch data see Den Besten 1987a, Ponelis 1993: 30-33, Roberge 1995: 79-81). According to Den Besten the fact that the Khoe Dutch examples generally show SOV (rather than SVO which is typical for pidgin/creole languages) could be a result of substratum influence since SOV is also the basic word order of the Khoe languages.\(^{22}\) At the same time surface SOV is the structure of Dutch foreigner talk. Early Khoe Dutch also makes use of a verbal marker -um/ -om, which might be a generalisation of the Khoe (Korana) noun suffix -m- (Den Besten 1987a: 35; see Valkhoff 1966: 219-220 for an alternative interpretation regarding the origin of the morpheme; note that the verbal suffix is particularly prominent in Kolbe's collection of Khoe sentences, where it might have been employed as a linguistic stereotype). The corpus of early Khoe and slave Dutch also includes several examples which are fairly close to continental Dutch, such as example (4) supposedly uttered by the Khoe Lubbert when five 'swarte jongens' came to his settlement trying to steal livestock. Example (5) comes from the evidence given by a slave in a court case.

\(^{22}\) In the early years of the colony about eleven (closely related) Cape Khoe varieties were spoken, most of which became extinct by the middle of the 18th century following the disintegration of Khoe society. In the 19th century Nama, Korana (!Ora) and Gri (Xri) were still spoken by Khoe groups in the Richtersveld and along the Orange River. Today the Khoe languages are represented only by speakers of Nama (Richtersveld, northern Cape, Namibia; for details see Traill 1995). In assessing possible substratum influence linguistic information available for Korana and Nama is used and generally seen as reliable (Den Besten 1986: 200-201, Holm 1988/89 II: 348).
In example (4) we find differentiation between subject and object pronouns for \textit{je/jou}, but the plural object pronoun \textit{ons} is still prominent in subject position. The infinitive is marked according to the norms of Dutch by the inflectional ending \textit{-en}, and even sentence brace occurs (\textit{ons sel ... geeven}). In example (5) we find an inflected present perfect (\textit{heeft gezegd}), and also two inflected infinitives (\textit{loopen, manqueeren}).

To assess the reliability of these fragmentary sources is difficult. In his discussion of early \textit{Negerhollands} data, Stein (1995) has suggested that creole utterances transcribed by matrilectal (or proficient L2) speakers of the superstrate are often fairly close to the lexifier language, while the actual spoken language appears to have been more divergent. A possible reason for this discrepancy is that the orthographical conventions available to the scribes did not have accepted ways of representing non-European substrate or innovative features. Writers thus wrote down only what they understood (i.e. what related to their own linguistic and orthographic system), and what they thought their readers would understand. Non-European features were frequently omitted and the closeness to Dutch evident in examples (3) and (4) could, thus, be the result of scribal editorship. At the same time we cannot exclude the possibility that these early representations of Khoe and slave Dutch reflect stereotypical perceptions of the nature of non-matrilectal Dutch as perceived by the court scribes (i.e. rather than omitting certain features, scribes might have exaggerated them).

According to Den Besten, the early Khoe Dutch data suggest the existence of a relatively stable Khoe pidgin (characterised by features such as the use of nominative \textit{ons}, the article \textit{die}, the verbal marker \textit{-um/ -om} etc.), which formed the basis for the development of a Dutch pidgin/creole among the slaves (1987a: 20-24, 1989: 222). Ponelis (1993: 33-34, 1995: 221) has rejected this hypothesis on sociohistorical grounds, i.e. the Khoe substratum was too small and not sufficiently integrated into the colonial society as to have served as a model for the development of a slave pidgin/creole. Our demographic information regarding the Khoe substrate
is unfortunately extremely fragmentary, and the actual size of the substrate community is almost impossible to assess. According to Elphick (1977: 23) between 4000 and 8000 Khoe lived on the Cape Peninsula during the early years of the settlement. However, far fewer were in regular contact with the colonists and their slaves. Ponelis (1993: 10) has estimated that only about 80 detribalised Khoe lived at the company’s fort in the 1660s, but from the early 18th century many more Khoe were employed by colonists (Shell 1994: 26-27). By the late 18th century (1780) about 20,000 (largely detribalised) Khoe and about 3000 San lived within the borders of the colony (Worden 1985: 11), i.e. Khoe constituted roughly 50% of the colony’s population (see table 2.2). The percentage of Khoe was particularly high in the eastern (and northern) districts where ownership of slaves was rare, while in the western Cape slaves outnumbered the Khoe.23 In other words, although our demographic information is sketchy, the evidence shows that in numerical terms the Khoe substrate was certainly not insignificant, although it was most dominant in areas where there were few slaves (i.e. the eastern frontier). Roberge (1995: 78) has furthermore pointed to a remark made by Mentzel (1921 [1785]: 49) for the early 18th century, in which Mentzel described regular encounters which took place between Khoe and slaves, and Roberge correctly argues that '[t]he degree to which a Khoe Dutch jargon would have stabilised into a pidgin would depend on the intimacy of their linguistic encounters with slaves’, rather than on their numerical strength.24

Den Besten’s argument for continuity between the early Khoe pidgin and later Khoe/slave pidgin/creole is based primarily on the existence of linguistic similarities between the slave data and earlier Khoe data, i.e. both are SOV, use the object pronoun ons in subject position, use die as a definite article and show examples of lexical continuity (tecken, i.e. ‘take’ Khoe loan from English, and kammene, i.e. ‘cannot’ attested for Khoe pidgin Dutch, see Den Besten 1987a: 20-24). Den Besten further argues that within the Afro-Asian substrate the initial pidgin was

23 The tax rolls for Stellenbosch/Drakenstein in the western Cape give the following numbers for 1806: 4,952 settlers, 3,922 Khoe, and 9,002 slaves (Worden 1985: 11), while the census from 1798 indicates a Khoe-settler ratio of 3:1 for the eastern districts (Swellendam/Graaf-Reinet, Ponelis 1993: 40, see also Shell 1994: 147). In Cape Town, on the other hand, the influence of Khoe was negligible: in 1821 38% of the Cape Town population were settlers, 35% slaves and 2% Khoe (the rest were free blacks and members of the Garrison; Shell 1994: 143).

24 Roberge, following Whinnom’s (1971) model, distinguishes between secondary and tertiary hybridisation. Bilingual contact (with untutored SLA) leads to the emergence of variable secondary hybrids (such as cocoliche spoken by Italian immigrants in Argentina), which can develop into (stable) pidgins when a third party acquires the developing, unstable L2 variety without having sufficient access to the original target (this then constitutes tertiary hybridisation). Thus, while Khoe Dutch was merely an unstable, highly variable jargon (or L2 variety), the ‘motor’ of tertiary hybridisation (and thus pidginisation) was possibly provided by the slaves. The result was a stable pidgin variety spoken within the Afro-Asian substrate; however, it was not nativised ‘in the sense that it provided the primary input for L1 acquisition; there was after all no actual withdrawal of the superstrate language’ (Roberge 1998: 21).
natived into a creole (Proto-Afrikaans I, from about 1680), while the Dutch used by the European settlers (although under the influence of the creolised variety) developed independently (Proto Afrikaans II). These two strands merged in the late 18th century into a koiné, resulting in the integration of b_ACL features into the varieties spoken by the European settlers: 'two types of Dutch, European and Pidgin Dutch, gradually coalesced to yield Afrikaans and its dialects' (Den Besten 1989: 234; this scenario elaborates Franken's 1953: 193, 202-203 and Valkhoff's 1966: 207ff. earlier discussions).\(^{25}\)

\[\text{English/ Dutch trade jargon} \quad \text{c. 1640} \]

\[\text{Dutch} \quad \text{\textarrow{v}} \quad \text{Khoe Dutch} \quad \text{\textarrow{v}} \quad \text{Slave Dutch} \quad \text{c. 1660} \]

\[\text{\textarrow{v}} \quad \text{Creole Dutch} \quad \text{c. 1700} \]

\[\text{(Proto-Afrikaans I)} \]

\[\text{Cape Dutch} \quad \text{c. 1740} \]

\[\text{\textarrow{v}} \quad \text{Proto-Afrikaans II} \quad \text{c. 1775} \]

\[\text{Afrikaans koiné} \quad \text{c. 1850} \]

Figure 2.5 Emergence of Afrikaans according to Den Besten's 'Convergence Model' (adapted from Den Besten 1989: 226). Top-to-Bottom arrows indicate diachronic development, horizontal arrows indicate linguistic influences.

Historiographical research has however shown that racial and social stratification increased as the eighteenth century progressed, which makes a general merger of the two varieties unlikely.\(^{26}\) Demarcational needs linked to questions of social status would have not only

\(^{25}\) Valkhoff (1966: 231) termed this the 'idea of a double origin' of Afrikaans, and summarised it aptly: 'Hence, in the history of Afrikaans, too, it was not always either Dutch or Creole, but the two linguistic currents may well have met and the latter may have stimulated the development of the former.' To postulate multiple ancestors for Afrikaans opens the question whether Afrikaans can or should be classified as a descendant of Dutch (Thomason & Kaufman 1988: 10), and accordingly Den Besten (1989: 245) questioned the classification of Afrikaans as West-Germanic despite the fact that transmission was (at least within the superstrate) 'normal'. However, as indicated at the beginning of this section, not too much importance should be assigned to such questions of classification (see also Epilogue).

\(^{26}\) There is some anecdotal evidence of the acquisition of b_ACL features by European children who acquired creolised varieties from slave children or their nannies (Van Rheede 1685, discussed in Scholtz 1963: 257ff.; for the early 19th century see Worden, Van Heyningen & Bickford-Smith 1998: 127). Mentzel (1925 [1785]: 108), however, denied that it was common practice among the colonists to
prevented to some extent the incorporation of possibly stigmatised pidgin/creoles features into acrolectal varieties close to Dutch, but also supported the emergence of sociolects (both within the free population and the Khoe/slave communities) thus complicating Den Besten's rather monolithic model (on the sociohistorical background see Guelke 1989: 82-84, Armstrong & Worden 1989, Worden 1985, Shell 1994). Roberge (1995: 81-82) thus argued that the degree of convergence between acrolectal and basilectal varieties would have been a function of the social standing as well as the geographical location of the speakers.

Several factors would determine the degree of influence of one code on the other. Members of the prosperous burgher class in Cape Town and the wealthiest wine and grain farmers in the Boland would not have known more than a smattering of the pidgin (if any). Settlers of lesser means in the Boland and along the frontier had at least a passive - often active - knowledge of the Cape Dutch pidgin. Acrolectal Dutch would have become more and more diluted with increasing social and geographic distance from centres of power. The extent of dilution would naturally be greatest in the rural areas along the frontier (where speakers were simply not as familiar with prestige norms), within the slave community generally by virtue of multiple inputs, and among the Khoe and Bastards. In other words, the Dutch language at the Cape of Good Hope formed a continuum from the most creole-like variety within the Afro-Asian substrate to an uncreolised extra-territorial variety of the European superstrate. The speech of individuals took on or avoided pidgin features depending on the interlocutor, the nature of their communicative networks, and the sociolinguistic circumstances (code-switching).

Assuming a continuum of 'creoleness' (in the sense of more or less contact-induced restructuring) is suggestive of the concept of the (post) creole continuum originally introduced

leave their children in the company of slaves (or Khoe): 'Kolbe's statement with regard to the bad upbringing of South African children is very misleading. It can apply only to the children of very poor or of dissolute parents. The average child is not allowed to run wild among the slave-children. On the contrary, he is tenderly cared for by his mother.' Although the use of slave-women as wet nurses was common practice, these were assigned a special role the household and had 'to do everything under the supervision of her mistress' (ibid.; on the practice of wet nurses and slave nannies see Shell 1994: 304ff.).

Shell not only emphasises social stratification within the slave community (1994: 49-58), but also remarked with regard to the colonists: 'There is [...] no evidence whatsoever for the popular notion of an early colonial "herrenvolk democracy" among a "master race" of European settlers. This golden age mythology found in the plaasroman (farm novel) of the 1930s could only have been latter-day compensation for the glaring inequality among early European settlers. Cape society perfectly illustrated Vilfredo Pareto's iron law: no matter how evenly goods are distributed initially, after a single generation ownership will be concentrated' (1994: 153). Note, that social standing was not necessarily a function of 'race' or 'colour'. Shell (1994: 23), for example, discusses in detail the low social status of knechts (wage labour of European descent), who constituted between 10-30% of the free population (1700-1750). Knechts were often excluded from the European society and thus 'drawn to slaves for companionship'. On the other hand, there were the so-called 'free blacks' who originated from the slave population, political exiles from Batavia and Chinese convicts, some of whom had gained considerable social standing (see Davids 1991: 32ff., Shell 1994: 232).
by DeCamp (1971a), and Ponelis' (1988: 126) description of Afrikaans as a 'behoudende akrolektiese kreoel' suggests such a line of thinking (see also Makhudu 1984: 96). According to DeCamp's original discussion, a post-creole continuum emerges as the result of a historical process in which the original bisystemic opposition of basilect (pidgin/creole; the variety most removed from the lexifier language) and acrolect (varieties close to the superstrate language, usually spoken by the settlers and a few private slaves) develops into a linguistic continuum via the emergence of intermediate varieties (mesolects which are thus 'younger' than basi- and acrolect). The process has been termed 'decreolisation' i.e. linear language change of the individuals' basilects into the direction of the acrolect thus giving rise to a variety of mesolectal forms. However, from early on linguists have argued that the linguistic continuum should not be seen as the result of a historical process but that it existed from the beginning of the contact situation.  

This is also the position of Rickford (1983) who proposed a quantitative model of decreolisation which takes the speech community rather than the individual's language as the starting point. Rickford has argued that the primary impact of decreolisation lies in the declining proportion of people who speak the basilect, and not in a decline in the 'purity' or 'creoleness' of the variety itself. In other words, during decreolisation individuals acquire (additional) lects closer to the acrolect without necessarily losing their knowledge of basilectal forms (see also Escure 1997: 64-65). In scenarios involving targeted SLA, on the other hand, we find replacement rather than extension of the individual's repertoire, i.e. the language learner usually replaces his 'basilang' with a 'mesolang' (and finally an 'acrolang') with no incentive to maintain the developmentally lower varieties once a more perfect approximation of the target has been mastered.

---

28 See, for example, Alleyne (1971: 170, 173-174; 1980: 182-183), Mufwene (1988, 1996: 101, fn.13), and Escure (1997: 61-63). Whether the creole continuum should be seen as uni- or multidimensional (see Le Page & Tabouret-Keller 1985, Rickford 1987a) is not of importance here. While a unidimensional model in which variables are ordered linearly in terms of 'creoleness' or 'standardness' is adequate to grasp the main structure of many contact situations, it will often interact with more complex patterns of variation (such as rural/urban, socioeconomic class, age, etc.).

29 Rickford's analysis of decreolisation is more than a simple re-interpretation of DeCamp's original model; as Mufwene (1988: 114 ) has emphasised 'Structural/formal change and change in the proportion of speakers using linguistic features "ain't the same thing!" If decreolization means the first kind of change, it cannot mean the second.'
Table 2.3 A quantitative model of decreolisation involving shifts in the percentage of speakers using the different lects (the percentages in each row do not add up to 100% as each speaker is believed to control a minimum of 2 lects; adapted from Rickford 1983: 300).

<table>
<thead>
<tr>
<th>Stage</th>
<th>Basilect</th>
<th>Lower-Mesolect</th>
<th>Mid-Mesolect</th>
<th>Upper-Mesolect</th>
<th>Acrolect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 1</td>
<td>33%</td>
<td>27%</td>
<td>40%</td>
<td>27%</td>
<td>27%</td>
</tr>
<tr>
<td>Stage 2</td>
<td>33%</td>
<td>33%</td>
<td>47%</td>
<td>33%</td>
<td>33%</td>
</tr>
<tr>
<td>Stage 3</td>
<td>13%</td>
<td>40%</td>
<td>40%</td>
<td>40%</td>
<td>47%</td>
</tr>
<tr>
<td>Stage 4</td>
<td>7%</td>
<td>47%</td>
<td>40%</td>
<td>47%</td>
<td>47%</td>
</tr>
</tbody>
</table>

While Roberge's (1998: 21) assessment that there is 'no evidence whatsoever to substantiate the view that decreolisation (loss of the basilect) represents a developmental stage in the history of Afrikaans' probably refers to DeCamp's standard model (i.e. decreolisation as 'de-basilectalisation'), societal decreolisation (i.e. the decrease of the number of speakers with knowledge of the basilect) as described by Rickford is a possible scenario. Thus, while the social conditions of the homestead society were conducive to the acquisition of mesolectal varieties closer to Dutch, basilectal varieties (which were acquired earlier from other substrate speakers, such as the Khoe in Den Besten's model) remained important, and possibly related to the solidarity function of language (especially amongst the disenfranchised).  

In this context attention should also be paid to Bickerton's (1995) more recent renunciation of the traditional model of decreolisation. Bickerton argues that new evidence (as, for example, presented in Baker & Corne 1982) forces us to reconsider our understanding of the creole continuum, and he suggests that,

wherever a creole continuum existed, it was likely to have developed 'backwards' rather than 'forwards'. That is to say, rather than the basilect forming first, then the mesolect, as the conventional picture assumed, the acrolect would have formed first, then the mesolect, and the basilect last of all. (1995: 317)

Bickerton's argument is mainly demographic, i.e. the initial predominance of Europeans in the colonies meant that the early slaves had, during the first few years, sufficient access to the lexifier language (see figure 2.4), and thus acquired L2 varieties fairly close to the dominant European language (acrolectal or at least mesolectal varieties). Such varieties persisted in later years when these acrolectal or mesolectal varieties became a form of self-identification for higher ranks of slaves (usually locally born slaves). Massive slave imports (the plantation stage)

---

30 Rickford's work in Guyana shows that basilects are often maintained after more mesolectal (or even acrolectal) varieties have been acquired. Particularly interesting is Rickford's informant Reefer, a cane-cutter, who initially appeared to be a monolingual basilect speaker, but was later shown to have also good command of mesolectal and acrolectal norms (Rickford 1983).
which reduced earlier (house) slaves and Europeans to a small minority then created the conditions for the emergence of basilectal varieties - and the continuum is complete. Considering the demographic conditions at the Cape (no large slave holder, slave-settler ratio roughly equal), Bickerton's revised model of creole development would account for the absence of typical pidgin/creole features (such as, for example, TMA systems) even in the most basilectal varieties CDV, which, on an 'absolute' scale of creoleness, seem to represent (lower) mesolects.

2.2.2 Creolisation or Interlectal Restructuring? - The Question of Stability

While today Afrikaans linguists generally agree that a contact variety spoken by the non-European section of the population has played a role in the development of Afrikaans, positions diverge regarding the actual type and intensity of the contact situation, and particularly the precise linguistic nature of the contact variety. In a recent assessment of the current discussion Roberge (1994b, 1995) distinguished two main positions: the 'creolists' and the 'interlectalists'.

- The 'creolists' assume that pidginisation/creolisation took place within the Afro-Asian substrate (Khoe, slaves) where it led to the emergence of a stable pidgin/creole ('there was a pidgin phase in South Africa - but not among the whites', Den Besten 1986: 185). Radical restructuring was prevented as access to the superstrate continued and the resulting pidgin/creole was much closer to the superstrate language than typical plantation creoles which evolved in almost complete isolation from the superstrate. The importance of the pidgin/creole for intergroup communication was restricted as within the slave community Creole Portuguese and Pasar Malay were used as lingua francas until the 19th century. Thus, demographically the pidgin/creole speaking community was small and its influence on the superstrate varieties spoken by the European colonists was limited (this position surfaces in the work by Hesseling (1923), Franken (1953) and Valkhoff (1966); it is most clearly articulated by Den Besten 1987b: 87, 1989: 227; and with some reservations also by Roberge 1994b, 1995, 1998).31

- The 'interlectalists' argue that the contact situation at the Cape led to untutored second language acquisition (SLA) of Dutch within the Afro-Asian substrate (Khoe and slaves). The result of this process was a continuum of interlanguages characterised by varying degrees of simplification and morphological reduction, hybridisation and substrate transfer, which served first a secondary languages and later as primary languages in the substrate communities. The rapid pace of language change as well as the existence of some non-Dutch features in Afrikaans

31 Within this position there is disagreement about which community formed the primary substrate of the developing pidgin/creole: the Malay and Low Portuguese speaking slaves (Hesseling, Valkhoff) or the Khoe (Den Besten).

Roberge has argued that the central issue dividing 'interlectalists' and 'creolists' is the question of 'what type of code was nativised within the Afro-Asian substrate: a stable pidgin or unstable, transient interlanguages emanating from gradient degrees of multilingualism and language shift' (1994b: 48, 1995: 77). According to Roberge's assessment the interlectalist position suffers from what he has called the 'conventionalisation problem' (1995: 77; also Roberge 1994b: 35f./1998: 15-16). In other words, assuming a dynamic interlanguage continuum comprising a large number of idiosyncratic and unstable varieties which arise spontaneously to cope with the communicative demands of the situation, makes it difficult to account for lasting linguistic influence on the acrolectal norms (on the general issue of stabilisation or 'crystallisation' see Thomason 1997: 82, on relative stabilisation as a necessary condition for interference and borrowing see Bailey 1971: 342). The problem with the interlectalist position is, thus, not their reference to processes of SLA, but their insistence that the resulting L2 varieties should be understood as individual solutions to communication problems, leading to a dynamic L2 continuum (see, for example, Ponelis 1993: 20, Van Rensburg 1994).

That pidgins/creoles share many features with cases of untutored L2 acquisition has long been acknowledged, and interlanguages have been shown to undergo 'fossilisation' (i.e. stabilisation before full proficiency is reached) under certain circumstances. This 'getting stuck' at some point of the language learners continuum does not depend on exposure to the target language alone but is the result of a 'complex function of ability and necessity' (Sankoff 1980a: 143), is illustrated by the existence of relatively stable foreign workers varieties, such as

---


33 A pidgin, on the other hand, has 'socially accepted norms' and is thus 'qualitatively different [...] from either jargons or interlanguages' (Roberge 1995: 70, also Thomason & Kaufman 1988: 168-169). However, Roberge (1994b: 36) acknowledges that if one assumes that first generation adult interlanguages stabilised through nativisation by children (which, however, is explicitly ruled out by Van Rensburg 1989: 138), then it is unclear to what extent the difference between the positions of Van Rensburg [...] and Den Besten [...] is not merely one of terminology'.

34 On the similarities between creolisation and untutored SLA see the overview given in Romaine (1988: 204-255). Note, however, that Bickerton (1981: 238) maintained that the two are fundamentally different processes. On 'fossilisation' see Selinker (1972) and Platt (1991) who has argue that Singaporean English is an example of such a 'fossilised interlanguage' which was then used as lingua franca.
Gastarbeiterdeutsch, where access to the target language is theoretically given as immigrant workers do not live in strict isolation from the native society. However, since the social interaction outside the work sphere is generally limited to the native immigrant community (at least for first generation immigrants), motivations for improving one's proficiency are generally low and a defective version of the target language is functionally adequate for the communicative demands in the work place (usually manual work in the production process) and the occasional encounter with German bureaucracy (Heidelberger Forschungsprojekt 1975: 44, Blackshire-Belay 1993). Mesthrie (1992: 185) has pointed out that an important characteristic of creole-like SLA is not only that it is untutored, but also that it happens in groups and leads to language shift. It is in such cases where the stabilisation of an L2 variety can be described as a group phenomenon that the difference between creolisation and SLA becomes academic.

Another contribution which touches on the relationship between SLA and creolisation comes from Baker (1990, 1997) who has questioned the validity of the concept 'target language' for pidgin/creole genesis, and who has argued that the target towards which acquisition is oriented is not the lexifier (as implicitly assumed by most of those supporting the 'interlectal' position), but what he calls the 'medium for interethnic communication' (MIC), that is, the emerging contact variety itself.

What I am suggesting is that the participants created a new language, suited for their immediate interethnic needs, and that they subsequently expanded and adapted this as their growing or changing needs demanded, drawing at all times on the resources available. These resources would include all languages known to participants, substrate and superstrate [...] as well as something akin to Bickerton's (1981) bioprogram, while the relative availability of these resources would be related to social factors. [...] (Baker 1990: 111)

In other words, the pressures of effective communication frequently override attempts to move closer to the lexifier language. While most SLA is targeted, untutored SLA is often only partially targeted and resembles creolisation in that proficiency in the acrolect is not necessarily aimed at. In other words, in both creolisation and untutored SLA the target is not necessarily the superstrate language, but the need to communicate.

Insisting on the necessary existence of a relatively stable contact variety to account for

35 The truly multilingual nature of the Cape household was described by Sparrman (1785: 58): 'At mealtimes, various European dialects, together with the languages used in commerce with the Indians, viz. the Malay, and a very bad kind of Portuguese were spoken at one time, so that the confusion was almost equal to that of the tower of Babel.' Baker's account falls, however, victim to the 'cafeteria principle', i.e. the idea that pidgin/creoles are mere mixtures of various rules and other items. It is questionable whether this is actually a problem, or not rather a challenge for further research which should try to establish general principles (such as markedness) which influence the selection of linguistic material (see Mufwene 1996: 104ff.).
substrate structures in modern Afrikaans, should not be misunderstood as implying homogeneity. It is generally acknowledged that pidgin/creoles, although crystallised to some degree, are highly variable linguistic codes (see Escure 1997: 41, 58ff). This complex reality between variability and relative stability which is characteristic of contact varieties was described by Schuchardt in his *Kreolische Studien IV* (1883) with reference to the situation of Malay-Spanish creole.

Dieses español de cocina [...] besitzt für den Sprachforscher ein großes Interesse. Es ist kein fertiges Patois, wie das Portugiesische von Macao oder Malacca, es existiert in den mannigfachsten Abstufungen, mit größerer oder geringerer Annäherung an die spanische Grammatik, mit größerer oder geringerer Verwendung malayischer Wörter, aber es ist kein immer wiederholtes individuelles Radebrechen mehr, das zwischen vielen Indem, mit Hintersetzung der angestammten Sprache, das regelmäßige Unterhaltungsmittel bildet; es besteht eine Art Überlieferung, es macht sich ein breiter Durchschnitt bemerkbar; die Spanier müssen sich an das "Küchenspanisch" gewöhnen und sich ihrerseits zu einer vermittelnden Sprechweise bequemen. (quoted in Bechert & Wildgen 1991: 129)

Before turning to the discussion of the 'mechanisms' of language change at the Cape, a few comments on the question of dating the emergence and stabilisation of a basilectal variety are appropriate. While the standard model of pidgin/creole genesis (which underlies Den Besten's model) assumes that contact languages emerge abruptly and acquire their distinctive features very quickly (usually within one generation; see Hancock 1987: 265), more recent research has shown that the structural development of contact languages might well be a slow and gradual process, which, for example, in the case of Mauritian creole stretched over 150 years (Baker 1997, see also Singler 1986, 1995, Arends 1993). Considering that the percentage of locally born slaves reached the 50% mark only in the 1760s, as well as the fact that large-scale shift from Khoi to Afrikaans started only around 1730 with some groups remaining bilingual until the late 19th century (the same was true for slaves speaking Creole Portuguese and Malay, see Valkhoff 1966: 261-264), stabilisation and development of the contact varieties might well have been gradual (see Arends 1995 on the relationship of slow nativisation of the slave community and gradual creole development). The continuing bilingualism of the substrate communities also means that substrate transfer could have surfaced considerably later than 1700 (which is acknowledged by Den Besten 1989: 227). Valkhoff's (1966: 217) original suggestion that 'after an initial "taking shape" Afrikaans developed in successive waves', can be interpreted from a gradualist perspective which acknowledges the possibility of a continuing and gradual 'basilectalisation process' understood as 'the consolidation of basilectal features into clearly identifiable sociolects, without ruling out the gradual introduction and integration of new alternatives' (Mufwene 1996: 101-102, 104).
2.2.3 The Mechanisms of Language Change at the Cape

Taking the sociohistorical and linguistic evidence together the consensus is that Afrikaans is an (almost prototypical) example of a semi-creole, i.e. a language which shows creole-like features and developed in a contact situation leading to moderate grammatical restructuring without necessarily descending directly from a creole ancestor (Markey 1982: 201-202, Thomason & Kaufman 1988: 251-56, Holm 1988/89 II: 339-40, Trudgill 1989: 232, Den Besten 1989, Bryn & Veenstra 1993: 30, Roberge 1994c). Contact is believed to have led not only to the acceleration of the rate of change and certain processes of simplification, but has also given rise to the development of linguistic characteristics unattested in the parent language group (i.e. West Germanic). Much discussion of the results of language contact at the Cape has focused on what is termed substrate transfer. Non-superstrate features which occur in the contact variety are believed to be traceable directly to substrate structures, which were transferred into the basilectal and later borrowed into the settlers' vernacular. In other words, we have a double-nested transfer: first from the substrate to the basilectal contact variety, then (via borrowing) into the acrolectal variety. Candidates for direct substrate transfer are objective vir (from Low Portuguese, Raidt 1983: 183-187), the reduplication of adverbs, adjectives, nouns, verbs and numerals (from Malay, Raidt 1983: 171-172), the double negation (from Khoe, Nienaber 1965/1994b, Den Besten 1986), the associative plural N-hulle (from Khoe, Nienaber 1994c), and the use of the irrealis marker kamma which occurs in non-standard Afrikaans (from Khoe, Nienaber 1963: 373).

Whinnom (1971) has argued that direct substrate transfer is typical for bilingual contact situations, while multilingual situations give rise to what he termed 'tertiary' hybridisations, i.e. not language mixture through borrowing, but convergent hybridisation, leading to structures which cannot be traced back unambiguously to either the substrate or the superstrate (see fn. 24). That assuming convergent hybridisation, rather than direct substrate or superstrate origin, might help us to untangle some of the linguistic puzzles in the history of Afrikaans has been the working hypothesis underlying some of Roberge's work. Regarding, for example, the Afrikaans verbal hendiadys (a complex double verb construction expressing progressive aspect, as in hy loop en sing), Roberge (1994c) suggests that serialisation of Dutch lexical verbs which is still attested for Griqua Afrikaans and some varieties of Western Cape Afrikaans, was common in basilectal varieties of CDV. The serial verb structure was re-analysed by acrolectal speakers in terms of their knowledge of Netherlandic structures (the early modern hendiadys construction with en), which thus facilitated the incorporation of certain aspects of the basilectal construction into the acrolect (i.e. optional affixation of the past participle marker ge- to the propositional verb, the extension of the construction to environments following finite modal and future auxiliary verbs, and the additional use of loop as a hendiadys trigger). In other words, basilectal features fused in such a way with inherited structures, that the resulting 'new rule' cannot be
traced back directly to either one of the different inputs. It is precisely this 'multilevel syncretism'
which, according to Roberge, characterises semi-creoles (and which at the Cape complemented
cultural syncretism as described by social historians).

This [...] seems to lie at the heart of what I understand a semi-creole to be. Measured against
'true' creoles, semi-creoles lack the massive restructuring of the grammatical systems. What we
find instead is a cycle of incremental innovation, accommodation, and integration [...]. The end
result is often hybrids that are etymologically opaque to the extent that they are not directly
relatable to any of their putative antecedents. (Roberge 1994c: 73).

Simplistic dichotomies between internally and externally motivated change are thus of
little use for the understanding of language change in a complex contact situation (on the danger
of such dichotomies see Dorian 1993, for a general discussion of the intersection of endogeny
and contact in creole formation see Mufwene 1996). At the same time, however, it is important
not to overemphasise the possible impact of contact-induced restructuring. Not all features which
are not directly attributable to the superstrate input are the result of substrate transfer or
convergent hybridisation, since endogenous innovation is common in extraterritorial languages. 36
Furthermore, large-scale morphological loss appears to be a general characteristic of
extraterritorial dialects. Thus, Reinecke (1937: 70-72) reported inflectional loss for the German
dialect spoken in Virginia, and Jersey Dutch (spoken in Bergen and Passaic Counties, New
Jersey), and Maher (1984) has argued that languages used in extraterritorial settlements (where
they came into contact with other languages) show striking similarities regarding the degree and
nature of their morphological reduction. Thus, Louisiana French, Canadian French and Virgin
Island French all show simplified pronominal systems, loss of inflectional endings in the verbal
system, loss of the auxiliary verb alternation in the perfect tenses (avoir vs. être), and a
predominance of periphrastic constructions over synthetic ones.

In sum, tracing the origins of Afrikaans involves (as Schneider 1993: 209 described for
AAVE) 'weighing data and interpreting the evidence in the light of conflicting theories and
possibilities offered to explain the outcome of a complex historical process', and four main
mechanisms of language change can be distinguished:

---

36 Lass (1997a: 174, 201ff.); see also Trudgill (1989: 228) on the increasing rate of endogenous
innovations in contact situations; for a general argument in favour of endogenous explanations in the
absence of clear evidence for contact see Lass & Wright (1986), Lass (1997a: 207-209).
(i) superstrate influence (direct continuation of superstrate features, but also evolutive tendencies of the superstrate language group)\textsuperscript{37}

(ii) substrate influence (Khoe, Creole Portuguese, Malay)

(iii) universal tendencies (general effects of imperfect learning amongst the substrate community such as morphological simplification)

(iv) multilevel syncretism (as discussed by Roberge)

In addition to these, the importance of mutual reinforcement has been emphasised recently not only in creolistics in general (see for example Schneider 1993, Siegel 1998) but also with regard to the history of Afrikaans in particular (already by Valkhoff 1966: 233). A good candidate for mutual reinforcement in Afrikaans is the loss of verbal inflection which was supported by general Germanic 'drift' and reinforced to differing degrees by processes of dialect levelling and the universals of untutored L2 acquisition (see § 6.1.1). As summarised by Jourdan (1991: 198) in a review article on developments in the field of pidgin/creole studies:

The emerging picture seems to be of an interaction among substrate influences, superstrate influences, and universal structures and faculties of language simplification. [...] these forces are not mutually exclusive but complementary and interactive; their relative weights vary from case to case.

2.3 The Insights of the Acrolect\textsuperscript{38}

While there has been a tendency in the study of language contact to focus on basilectal varieties, Rickford (1974) suggested in a paper titled The Insights of the Mesolect that 'rummaging in the mesolectal topsoil' could provide us with new data for a better understanding of the outcomes of language contact. More recently Escure (1997: 65) has shown that acrolects have equally been ignored in the study of contact-induced change, and are often mistakenly

\textsuperscript{37} As Roberge (1994b: 26, also Valkhoff 1966: 234) has pointed out, one needs to be careful when claiming continuation of linguistic features reported for dialects of the superstrate: 'Invoking dialect substrata in order to account for the unexplained residue requires that two conditions be satisfied. First, there must be a metropolitan dialect that provides a plausible linguistic model. Second, there must be independent evidence for the presence of speakers of the dialect in question 'at the relevant time and in sufficient numbers' [...]'. Note, that in this study the concept of superstrate influence is not limited to the direct continuation of superstrate features (including features reported for dialects of the superstrate), but also refers to general evolutive tendencies of the superstrate, including what one might want to term 'convergent evolution' within a language family, i.e. the independent development of similar features across a number of related dialects or languages.

\textsuperscript{38} The terminology of basi-, meso- and acrolect is commonly used for the description of language variation in multilingual societies (see, for example, Platt 1975, Mesthrie 1992), irrespective of the question whether a pidgin/creole exists or existed in the speech community. In such contexts the varieties structurally most different from the superstrate are usually identified as basilects.
equated with the metropolitan standard language (in this case continental Dutch). Acrolects, however, are better described as (generally highly variable) linguistic innovations characterised by the incorporation of linguistic features which have their origin in the contact situation itself. Unlike standard languages, acrolects usually have no overt set of linguistic norms and are primarily pragmatically motivated (i.e. depend on the formality of the situation). In other words, the concept of the acrolect is both absolute (on the level of the speech community) and relative (on the level of the individual).

An acrolect would, thus, be the highest variety an individual is able or willing to produce in the relevant formal contexts. [...] Indeed, one person's acrolect may be another's mesolect. (Escure 1997: 66, 74, emphasis in the original)

The mesolect, on the other hand, differs from acro- and basilects not only in terms of its linguistic structure (although the boundaries between acro-and mesolects are 'extremely difficult to establish', Escure 1997: 67), but also with regard to its social meaning (neither familiarity nor formality) and is typically used in semi-formal situations. The investigation of acrolectal and mesolectal varieties is of no less importance than the investigation of basilects, and equally important for our understanding of complex contact situations.

The European varieties collected by Van Oordt in the *Kaapse Taalargief* (§ 2.1) are probably best interpreted as examples of acrolectal usage. The same applies to documents such as the diary written by Johanna Duminy (Roberge 1994b: 60-63, 1995: 78-79, 1998; 24-26). The linguistic usage in the diary is close to metropolitan Dutch (e.g. preterite and pluperfect are still intact, the distinction between finite and non-finite, strong and weak verbs is preserved), but contains also many extraterritorial features (e.g. gender has almost disappeared and the verbal plural marker -e(n) is in the process of being replaced by singular forms). However, typical characteristics of Afrikaans such as object pronouns in subject position, the 'double' negation and the demonstratives *hierdie* and *daardie* are absent. In general, the 'formalizing effect' (Kytö & Rissanen 1983: 474) of writing probably skewed much of the historical sources in the direction of acrolectal and possibly (upper) mesolectal varieties.

Just as there is no reason to exclude acrolects from our study of the results of language contact, there is also no reason to exclude a priori any time period as irrelevant. As discussed in § 2.1 and §2.2.2 historical linguists have so far identified the 17th and the 18th centuries as central for the historical development of Afrikaans. However, the continuing variation in the historical record suggests that the emergence of Afrikaans might have been slower than originally assumed, shaped by different influences and differing socioeconomic conditions from the 17th to the late 19th century. Furthermore, certain important characteristics of Afrikaans (such as the 'double' negation, invariant *se* and the demonstratives *hierdie/daardie*) appear in the sources only
from the mid-19th century, which suggests that the 19th century was not necessarily a time of consolidation and stasis, but might have seen the emergence (or at least incorporation) of new features in meso- and acrolectal varieties. Researching the patterns of variation in the late 19th and early 20th century is of interest since at this time CDV was in an intense contact situation with two standard languages via the improving education system (English and standard Dutch), leading to bilingualism and possibly bidialectalism in the speech community. Sociohistorically the period under investigation was characterised by industrialisation and urbanisation, and thus, by the increasing development of weak ties which (according to Milroy & Milroy 1985b) are particularly important for the spread of new linguistic variants.
THE 19TH CENTURY AND THE 'DISCOVERY' OF THE VERNACULAR

Every time that the language question appears, in one mode or another, it signifies that a series of other problems are beginning to impose themselves: the formation and growth of the ruling class, the need to stabilise the most intimate and secure links between that ruling group and the popular national masses, that is, to reorganise cultural hegemony.

Antonio Gramsci Quaderni del Carcere a cura di Valentino Gerratana (Q29, 1935)

You get the illusion that those [entities] are just there and are being named as they exist. But they can be named quite differently and organized quite differently depending on the way the knife moves. [...] It is important to see the knife for what it is and not to be fooled into thinking that motorcycles or anything else are the way they are just because the knife happened to cut that way. It is important to concentrate on the knife itself.


3.1 The Rise of the Vernacular Writing Tradition

Linguistic descriptions of 19th century CDV have frequently focused on texts published in the periodical press which were written deliberately in what was meant to represent the 'vernacular' (see, for example, Lätti 1978, Uys 1983, Van Rensburg 1983, Van Rensburg & Combrink 1984, Ueckerman 1987, Raidt 1994 [1992]). Dialect-literature constitutes an important, albeit highly problematic source for sociohistorical analysis, and in order to assess the linguistic representativity of the texts (i.e. whether they can be taken to reflect actual patterns of communication and linguistic variation) the conditions of their production and reception need to be understood. In short the texts must be properly contextualised.

The first representation of CDV speech exists in an anonymous parody (Lied ter eere van de Swellendamsche en diverse andere helden bij de bloedige actie aan Muizenberg in dato 7 Aug. 1795) ridiculing the patriotic citizens of Swellendam (reprinted and discussed in Raidt 1991: 151-152). Over 30 years later M.D. Teenstra, who during 1825 spent four months at the Cape, included in his travelogue a dialogue portraying the colonists' spoken language (1830, reprinted in Nienaber 1971: 7-9). The text is accompanied by a short linguistic commentary in which Teenstra describes the dialogue as an attempt meant to introduce the reader to the most salient features of CDV ('Ziedaar in een kort bestek de grootste en misschien de meeste taalfouten en voor den vreemdeling meest ongewone gezegden en eigene spreekwijzen deser inwoners'), and Roberge thus characterised the dialogue as a literary artefact, a 'composite description of potential features that might have been found in any or all varieties' (1994a: 164). Travellers' anecdotes are generally difficult to verify, and 19th century travel accounts are notorious for their

1 Translation by Steinberg (1987: 206).
insufficient attention to the truthful representation of facts, focusing instead on the amusement of prospective audiences, and thus financial success (Batten 1978: 30, Brenner 1989: 14). However, since many of Teenstra's observations agree with what we know from other sources, the dialogue constitutes a nevertheless important secondary evidence for the history of Afrikaans.

From the 1820s short journalistic pieces written in what was meant to represent the 'vernacular' started appearing in the emerging Dutch press. One of the first Dutch periodicals at the Cape was *De Verzamelaar* (first issue 7.01.1826, later continued under the name *De Kaapsche Courant, Afrikaansche Berigter (of Verzamelaar)*), a satirical-political paper which has been described as 'a kind of Dutch *Punch*' (Meurant 1885: 75). *De Verzamelaar* was owned and edited by Joseph Suasso De Lima (1791-1855) who had arrived at the Cape in 1818 from the Netherlands. Using different humorous pseudonyms (such as Jan Hennepikker, Grietje Haanekeam, Griet Geldenaar) De Lima published several fictitious letters written in a highly variable, non-standard form of Dutch, in which he commented satirically on social and political events of local importance (reprinted in Nienaber 1971). While De Lima's pseudonymous characters were meant to represent members of the colonist class, Charles Etienne Boniface (1787-1853) in the 1830s published vernacular letters and dialogues, aimed at representing the speech of Khoe. Well known are his fictitious dialogues with the Khoe Hendrik Kok (published in 1830, for a linguistic analysis of these dialogues see Franken 1953: 188-207, Scholtz 1965: 76-103). The dialogues as well as a fictitious letter signed with the nom de plume 'Adam Slokker' were published in the *Zuid-Afrikaan* which was at the time under the editorship of Boniface. The texts are characterised by almost complete loss of verbal inflection, frequent use of the 'double' negation and a non-standard pronoun system (see § 8.1.3). CDV- speaking Khoe characters also occur in Boniface's popular comedy *De Nieuwe Ridderorde of De Temperantisten* (1979[1832]), while the European characters in the play speak Dutch. The stereotypical portrayal of Khoe-

---

2 However, factual inaccuracies in travel accounts are not necessarily conscious distortions, but can also be attributed to the possibility that the traveller finds himself in a situation which he cannot grasp intellectually. The interpretation of travel accounts thus raises important epistemological questions.

3 The anecdote narrated in the Hendrik Kok dialogues is based on an incident mentioned by John Philip in his book *Researches in South Africa* (1828). Philip reported that the magistrate W.M. Mackey from Somerset (now Somerset East) punished a Khoe, who had emptied a small cask of brandy belonging to the magistrate, not only with imprisonment, but kept him and his family for further three years under conditions of forced labour. However, Philip seemed to have distorted the incident and was sued for £1000 (for the historical background see Nienaber 1971: 43-44). The *Zuid-Afrikaan* was well known for its opposition to the negrophile position of the British humanitarians (Bosman 1930: 3), and Boniface's satirical reworking of the incident can be seen as reflecting the editorial bias of the paper.

4 Boniface, who was born in France, lived from the age of six in the Seychelles before arriving at the Cape in 1807. In the Seychelles a French based creole was well entrenched and widely spoken by the descendants of the slaves (Holm 1988/89 II: 401-403). Whether this experience influenced Boniface's perception and literary representation of the linguistic situation at the Cape is a matter of speculation, yet
characters as the 'jolly Hotnot', a quickwitted, irresponsible character whose merriment is typically alcohol-induced (Van der Merwe 1994: 22-23), is also evident in Andrew Geddes Bain's variety show piece Kaatje Kekkelbek, or Life among the Hottentots (1838, reprinted in Nienaber 1971) which enjoyed much success in the colony (Mesthrie 1993: 55). Unlike the texts by Boniface, however, Bain's version of vernacular speech is characterised by extensive code-switching between CDV and English (see § 8.2.1).

Neither De Lima nor Boniface or Bain provided glosses or linguistic commentaries (as Teenstra did for his European audiences) which indicates that their readers had no particular difficulty in understanding the texts, and were able to grasp their 'entertainment value' without requiring additional information (possibly knowing the dialect themselves as a resource to be called upon in certain situations). Following Winer's (1984: 192) discussion of vernacular representations of Trinidad Creole this can be seen as an indicator for the genuine character of the vernacular representations.

A little-known, transient, idiosyncratic or artificial variety would not have been comprehensible, nor would it have been amusing through its familiarity to the majority of readers. And if [...] the object of the pieces was to ridicule well-known local figures, and to make social and political commentary, it would have been necessary for readers to have understood the texts almost entirely.

There is, however, another possible interpretation, namely that the texts reflect a common perception of what constituted the 'vernacular' in a general undifferentiated sense, a kind of stylised or stereotypical 'General Low Colloquial', an amalgamation of different non-standard features which did not necessarily coexist (in general or in a given frequency) in the speech of any individual, but which endowed the texts with the (stereotypical) characteristics of rustic, nonstandard speech. In this sense vernacular or dialect writing resembles what Preston (1992) has called 'variety imitations', i.e. members of a given social group imitate the (generally highly marked) linguistic behaviour of another social group. Preston has argued that such imitations often become part of folk culture/popular culture, and that the ability to perform such imitations is a skill highly valued amongst those who engage in these culturally bound modes of story- and joke-telling. Dialect imitations are not necessarily faithful to the actual linguistic characteristics

biographical data of this kind should not be discounted. This also applies to anecdotal information about Boniface's personality which was reportedly given to exaggeration and extreme opinions (see Nienaber 1939).

5 See Ellis (1994: 118) for a similar position with respect to American dialect literature; see also Britto (1986: 314) for an interpretation of vernacular writing as 'stylised' and 'non-authentic', and Mufwene (1988: 113) on the highly regularised character of literary Gullah.
of the dialect, yet their lack of linguistic accuracy does not seem to impact on the comic effect of the imitation. Overgeneralisations of otherwise rare linguistic features are typical for dialect imitations (see Baugh 1992, Janda & Auger 1992: 209), and it was suggested by Ellis (1994: 139-140) that such overgeneralisations are most pronounced in cases where the dialect is used for comic or political purposes (see also Campell 1956, Brown 1995: 42 on 'enhancement of contrast' i.e. the exaggeration of differences as a typical feature of all stereotyping).

Interestingly the overgeneralisation of rare characteristics appears to have a cognitive basis. In social psychology this is usually discussed in connection with a phenomenon known as the 'illusionary correlation effect' which can be summarised as the 'erroneous perception of the co-occurrence of rare characteristics' (McGarty & De La Haye 1997: 145). In controlled experiments people were shown to associate regularly infrequent characteristics with equally infrequently occurring group membership (minority groups), although in reality no correlation existed (i.e. perceptions were found to be biased towards infrequent events).\(^6\) Regarding the explanation of the phenomenon Hamilton & Gifford (1976) originally interpreted it as a result of an enhanced perception of rare features, that is, people show a cognitive bias or sensitivity towards the perception of statistically infrequent events, which are more distinctive and thus trigger attention. The original distinctiveness-based hypothesis of Hamilton & Gifford was, however, challenged more recently by Fielder (1991, Fielder, Russer & Gramm 1993) who interpreted the phenomenon as a result of impaired rather than enhanced memory. In other words, estimating frequencies is generally a difficult task and extremely low frequencies although correctly perceived as lower than high frequencies, are typically perceived as being higher than they really are; the phenomenon is thus a result of information loss.

Louis Henri Meurant (1812-1893), who was acquainted with both De Lima and Boniface (Meurant 1885: 75) took up their earlier attempts at vernacular writing in the 1840s with his letters and dialogues published in *Het Kaapsche Grensblad* (Grahamstown; Meurant was the founder, owner and editor of the paper). Again the pieces were published either anonymously or under humorous pseudonyms, commenting on all kinds of local events and sociopolitical issues (see Scholtz 1965: 265). In 1860 Meurant published two vernacular dialogues in *The Cradock News* (he had left *Het Kaapsche Grensblad* in 1851), arguing for the political separation of the eastern and western Cape (the dialogues were reprinted as a booklet in 1861). The dialogues were widely read and the two fictitious characters, Klaas Waarzegger and Jan Twyfelaar, were soon integrated into the popular culture of the time (see below). From the 1860s onwards the production of vernacular letters and dialogues developed into a popular genre, and the stream of contributions (although fluctuating) increased steadily (see figure 3.1). There is little doubt that

---

\(^6\) Note that the concept of 'minority group' as reflected in the these experiments only refers to infrequently encountered group membership, but does not have any sociopolitical associations.
the aim of these writers was to write what was seen as 'the vernacular', and not Dutch (Scholtz 1965: 277). The popularity of dialect writing was not limited to the Cape colony, and similar texts also occurred in the periodical press in the northern republics (Scholtz 1964: 177f., Nienaber 1967).

Judging from figure 3.1, Van Rensburg & Jordaan (1995: 109) would seem to be wrong in their assessment that the end of the separatist movement also meant the end of the use of Afrikaans for socio-political commentary. It appears that it was not so much the content of Meurant's writings that mattered (i.e. the issue of separatism), but that through his writings the genre of vernacular letters and dialogues became established as a popular form of social commentary which was characterised by irony and spot, and a deliberately naïve and rustic style (referring, for example, to Queen Victoria as tante, see Nienaber 1968: 11-16), often interspersed with personal attacks. Vernacular letters were either published anonymously or using noms de plume some of which established prototypical characters (Arme Boer, Boerenseun, Een Burger), while others appear to have been part of a comic popular repertoire (Polly Lekkergoed, Man in die Maan, Koos Paapegai, Jantje zonder naam en zonder geld). Intertextual references are common (as in pseudonyms such as Maat van die Man in die Maan, and numerous references to Meurant's popular character Klaas Waarzegger, such as Gert and Piet Waarzegger, Klein Klaas Waarzegger, Klaasie, Klaas Zoon and Klaas Waarzegger jun., see Ueckerman 1987: 4-7, also Nienaber 1942: 15, for a list of pseudonyms see Nienaber 1967: 63-100). Most pseudonyms
characters also occurred. While the vernacular in the mouths of colonists became established as a form of entertaining social commentary, its association with Coloured characters was generally of a more burlesque type (see for example, Tooneelen uit het Politie Hof, written by adv. J.J. De Kock, published in Het Volksblad 1859-1863). Joyce's comments on 19th century English dialect literature fit the situation at the Cape rather well.

[A]uthorship is irrelevant: the uniformity of method, tone, and content [...] reflects intertextual influences, but far more than this it reflects a uniformity of expectation in their audience. To sell they had to correspond to the people's self-mythology, and in responding they also shaped that mythology. [...] Authors' nonne de plume established the right to speak for particular localities. (Joyce 1991: 163)

Using the vernacular in the increasingly popular genre of praatjes and boerenbrieven became a 'way of speaking', an alternative (if not subversive) cultural practice when taking the point of view of those located outside of the dominant (economic and political) power structures of society, and whose ideas differed from the orthodoxy of the dominant discourse. This is evident in the vernacular letters published under the pseudonym Jantje Eenvoudig (alias T.F. Burgers) which deal with the conflict between progressive and orthodox tendencies in the NGK (for the historical background see Raidt 1994 [1987], on T.F. Burgers see Marais 1993: 133-136). While the orthodox side published its position exclusively in Dutch, the liberal side (represented by Jantje Eenvoudig) utilised the vernacular (Raidt 1994 [1987]: 275, also Nienaber 1942: 8). A similar pattern is noticeable in the 1860s when Meurant published his Zamenspraak. In a letter to Robert Godlonton (dated 8/1/1861) Meurant explained his use of the 'vernacular' as a reaction to the conservative Dutch press in Cape Town which argued (using Dutch) against the proposals for political separation (quoted in Nienaber 1968: 6-7).

There has been a discussion in the literature as to whether the upsurge of vernacular publications around 1860 qualifies as a separate language movement (see Nienaber & Nienaber 1941: 73, Scholtz 1964: 181f., Du Plessis 1986: 15-27). However, attempts to identify separate language movements are ultimately futile since they misjudge the very nature of social movements, which typically blend into each other and are based on one another. Thus, rather than focusing on classificatory questions (i.e. what qualifies as a language movement and what are its boundaries), the interest should lie in understanding the interconnections of these language-centred activities (on the importance of relating new social practices to existing meanings and

---

7 Criticisms of the vernacular writing practice abound in the press, and many voiced discomfort in cases where the 'vernacular' was used to discuss topics which fell into the domain of high-culture such as education or church matters.
practices see Tarrow 1992: 189, also § 3.5). The protagonists of what became known as the first (1875-circa 1904) and second language movement (circa 1903-1914) for Afrikaans (see § 3.2) were well aware of their interconnections and continuities with the pre 1870s vernacular writing tradition (see Du Toit 1909 [1880]: 3ff., De Waal 1932: 2, also Nienaber 1942: 76, Scholtz, 1964: 182). That vernacular publications such as Meurant's Zamenspraak were deeply integrated into the popular culture of the time is evident from the following letter to the Zuid-Afrikaan.

Onder mijn schrijwe denk ek soo aan de brieue van 'Klaas Waarzegger' wat omtrent 12 jaar geleden in die koerante was. Die skrywer daarvan mot een verstandige kerel gewees het, want hij kan et nie beter aangeleg het, om te seh wat hij wou seh, dan in die taal van Afrika. Dinge wat kort gelede gebeur is het die mense al vergeet al, maar praat over 'Klaas Waarzegger' en dadelik begin hulle te lag; hulle kan hom nie vergeet nie. (Zuid Afrikaan 5/7/1873)

The fact that most vernacular texts used pseudonyms makes it difficult (if not impossible) to identify their authors. Those authors who have been identified belong to the middle classes and the intelligentsia, i.e. individuals who were in general perfectly able to write 'correct' Dutch and for whom the use of the vernacular was thus a conscious choice (see Nienaber 1942: 21). We know something about the targeted audience from the above-mentioned letter from Meurant to Godlonton, where Meurant identified 'Dutch Boers' as the targets of his propagandistic writings. There is little doubt that the category of 'Boer' is a wide one; however in the context of the separatist movement those whose support was needed and who Meurant tried to convince with his dialogues were likely to be those farmers to whom cheap labour, a local deeds registry office and a better infrastructure mattered, in short independent farmers with at least moderate property holdings (Le Cordeur 1981: 43, 284f.). Furthermore, the audience must have been familiar with the norms of Dutch to understand the linguistic humour of the texts, i.e. to understand that these texts were dialect imitations.

Thus, far from constituting an authentic representation of the language of 'the people', the dialect literature was produced and (probably) consumed by social groups who, although not politically or economically dominant, can be seen as peripheral members of the dominant culture. In other words, while not being part of the 'ruling class' they were not without (typically locally based) political and economic power (i.e. the middle classes including the propertyless intelligentsia; see § 4.3.2 for a more detailed discussion of social class and class analysis). The vernacular writing tradition should thus be understood as a socially mediated perception of what the vernacular is, rather than direct access to the spoken language. Raidt's assessment of the letters of Samuel Zwaartman (alias H.W.A. Cooper, a law agent and later advocate) as a reliable reflection of the 'destydse omgangstaal' (1968: 117, also Läti 1978) is thus debatable.

Considering what we know about the processes of standardisation in general, it is, furthermore, possible that the production of vernacular texts contributed to linguistic focusing,
or at least to the emergence of an increasingly focused typological conception of what constituted the 'vernacular' (or 'Afrikaans' as it came to be known). That is, people who decided to try their hand at this new genre had a model in earlier attempts, and in turn their efforts constituted a model for later attempts. One might term this (following Stein 1994) 'initial standardisation', i.e. certain selection decisions were taken but the process was not yet coordinated or institutionalised (nor for that matter necessarily aimed at vernacular elevation). Table 3.1 shows the relative frequencies for the Dutch relative pronoun die (vs. Afrikaans wat) for two successive time periods in the Transvaal periodical press. Dutch forms are clearly used much less in the later period, suggesting that forms marked as 'Afrikaans' gained popularity.

<table>
<thead>
<tr>
<th></th>
<th>no. of relative clauses</th>
<th>die as relative pronoun</th>
<th>% of die</th>
</tr>
</thead>
<tbody>
<tr>
<td>1866-1883</td>
<td>325</td>
<td>82</td>
<td>25.2</td>
</tr>
<tr>
<td>1884-1889</td>
<td>1770</td>
<td>220</td>
<td>12.4</td>
</tr>
</tbody>
</table>

Table 3.1 Wat and die as relative pronouns in the Transvaal periodical press 1866-1889 (from Van Rensburg & Combrink 1984: 115). Note that the difference between the two periods is statistically significant, $\chi^2$=36.08281, df=3, p<0.001.

The vernacular writing tradition was not limited to the European colonial community. Most notably the Muslim community at the Cape developed its own tradition of vernacular writing from the 1840s (Muller 1962, Davids 1991). Here the vernacular was used primarily to facilitate the teaching of religious practices since proficiency in Arabic was low at the Cape. At least 72 CDV texts written in Arabic script were published between 1869-1957 (Van Selms 1979: viii, Kähler 1971: 70-188, Davids 1991: 91-93). In addition a folk musical tradition, especially the so-called Ghommaliedjes (picnic or street songs based on Dutch folk songs), made use of Dutch as well as varieties of CDV (Du Plessis 1944: 57ff., Bickford-Smith 1995:

---

8 The term 'Afrikaans' occurred sporadically in the first half of the 19th century, but until the 1870s the dominant term was Kaqps(sch) Hollands(ch) (Viljoen 1896: 17). Even today some communities such as the Griqua use the term Hollands rather than Afrikaans when referring to Netherlandic varieties spoken around 1900. Janson (1991: 26) remarked in the context of Romance linguistics that the emergence of new language names is typically connected with the establishment of new written languages: 'the introduction of language names is closely connected with the creation of written language forms [...] the early history of names as well as the written forms is linked with powerful and numerically small groups in society rather than with people in general'. See also § 2.1, fn.5.

9 Davids (1991: 110ff.) indicates that Abu Bakr Effendi's Bayâmudîn (1869) came to constitute a model of how to write 'Arabic Afrikaans'. Earlier examples of Arabic Afrikaans (the so-called Koplesbooks) showed a much more variable language with strong influence from Dutch. Abu Bakr Effendi was a Turkish religious scholar whose primary language was English (Davids 1991: 117); with the publication of the Bayâmudîn Effendi tried to introduce Hanafite teachings to the Cape, which so far had followed the Shafîite school. The Bayâmudîn was initially circulated in manuscript form, and was printed only in 1877.
A tradition of vernacular writing also developed on the Moravian mission stations. In 1873 the conversion narrative *Benigna van Groenekloof, of Mambre* was published by the Genadendal press. The main text, narrating the story of Benigna's conversion to Christianity is written in Dutch, the dialogues however are written in CDV. The Moravian periodical *De Bode van Genadendal* (founded in 1859, published under the title *Die Huisvriend* from 1914) also published vernacular texts typologically similar to those known from other Dutch periodicals of the time, and offered social and political commentary in the guise of dialogues or fictitious letters (*Oom Gezels, Klaas Vertel aan Oom Koenraad, Vriend van Allen aan Oom Alspraat*; see also Belcher 1987).

In her assessment of the representations of Trinidadian creole in the *Spectator* texts Winer (1984: 194-196, also Winer 1993: 130-131) formulated two criteria for determining the reliability of literary vernacular representations:

(i) there should be internal consistency of the linguistic features within a given text, and
(ii) the representations should show external consistency with other evidence or contemporary language.

Comparing the vernacular texts (including the Muslim and the Moravian tradition) with each other and contemporary (standard) Afrikaans both criteria appear to be fulfilled especially with regard to morphology, and to a lesser degree syntax and phonology where more variation is noticeable (see Raidt 1994 [1992]). However, regarding the criterion of external consistency the situation is problematic when documents which are not deliberate and conscious vernacular representations (such as private letters and diaries) are included as an independent check.

10 Just as in the European colonial community, support for vernacular writing was not uncontroversial. In the first report on the educational system in the Cape Colony (1891) Abdol Burns a cap-proprietor and General Secretary for the Muslim community is quoted as follows: 'Do you want to see English and Dutch both taught? Yes; but not what they call Cape Dutch. I prefer proper Dutch. I do not talk what is called the 'taal' in my house. My son is able to read English, Dutch, and Arabic. He was taught English and he picked up the Dutch.' (First Report and Proceedings 1891: 96).

11 The Moravians on St. Thomas seem to have encouraged the use of Dutch creole (*Negerhollands*) for religious purposes (see Stein 1989, 1995), and also engaged in the compilation of dictionaries and grammatical descriptions. However, no indication of similar activities was found for the Moravian church at the Cape before the 1870s.

12 Kytö & Rissanen (1983) distinguish in their linguistic corpus between 'marked informal' and 'unmarked informal' on the basis of the degree of conscious manipulation by the author. 'In marked varieties, the writer's choice of the stylistic features which increase the impression of the formality or informality of his text are more conscious than in unmarked varieties' (1983: 475). Dialect literature can
Although the morphological and syntactic forms which occur in the vernacular texts are also present in the private documents (loss of inflections, loss of gender, objective *vir, ons* in subject position etc.), the difference in frequency is considerable. For example, while the (post 1860s) vernacular writing tradition shows almost complete loss of verbal inflections and loss of gender, these aspects are still highly variable in 19th century family letters and diaries. This is not to say that the vernacular writing tradition was not rooted in the actual patterns of language use, but the very fact that members of the intelligentsia (such as teachers, ministers and advocates) were at the forefront of this movement deserves closer attention, as does the possibility that as 'variety imitations' these texts might be characterised by stereotypical overgeneralisations. That dialect literature is often of limited use for identifying the quantitative aspects of linguistic variation was pointed out by Bailey & Ross (1992: 519): 'while literary dialect is often useful in identifying the presence or absence of features, it frequently treats quantitative features qualitatively'. In order to assess the status of the dialect literature, it is necessary to compare these texts with other types of contemporary linguistic evidence, such as the corpus of private papers discussed in chapters six to eight.

### 3.2 Afrikaner Nationalism and Early Vernacular Standardisation

Afrikaner nationalism with its romantic identification of language and nation was of central importance for the sociolinguistic development at the Cape where it led as elsewhere in the world to 'a golden age of vernacularising' (Anderson 1991: 71).\(^{13}\) Historians of nationalism have emphasised that 19th century nationalist movements were typically characterised by the 'invention' or '(re-)construction' of national characteristics and symbols, that is, of certain practices which symbolised the social cohesion of the in-group, and exaggerated differences with the out-group (see Hobsbawm 1983, Du Toit 1983, Gellner 1983: 55-56, Hofmeyer 1987, Fishman 1989 [1972]: 120-121). Although continuity is typically claimed with the remote past, few of the propagated national symbols are older than the movements themselves.

\(^{13}\) The identification of language and nation is usually dated to Herder and eighteenth century German romanticism, although similar thoughts were already articulated during the French enlightenment (specifically by Condillac, see the papers collected in Koepke 1990). By the second half of the 19th century the model of the independent nation state based on cultural and linguistic unity was a common place in European intellectual circles, a political/cultural model ready for 'pirating' (Anderson 1991: 81), and it can be assumed that young Afrikaners who went to study in Europe in the 19th century (see Scholtz 1964: 90, also Marais 1993) were introduced to these discussions. Furthermore, continuous European immigration (as well as administrative and political ties with the European powers) facilitated the influx of European social, political and scientific thinking.
We should not be misled by a curious, but understandable paradox: modern nations and all their impediments generally claim to be the opposite of novel, namely rooted in the remotest antiquity, and the opposite of constructed, namely human communities, so natural as to require no definition other than self-assertion. (Hobsbawm 1983: 14)

Acknowledging that nationalism and national symbols (including language) are creative constructs emphasises the importance of human agency and imagination, but this is not to say that national symbols are entirely or even primarily artificial. Nationalist movements typically work with what already exists (such as the vague and unspecified heritage of pre-modern ethnically based memories, myths and traditions), leading to manipulation and re-construction rather than downright invention (Smith 1993). Their activities can be likened to the work of the bricoleur, who is engaged in the novel combination and interpretation of existing elements, resulting in the formation of new structures (Lévi-Strauss 1976: 19-22, on the concept of bricolage see also Lass 1997: 313). Language, the hallmark of the romantic concept of a nation, did not escape the creative activities of nationalist movements, and Deutsch (1968: 603-604) has argued that from the mid 19th century such movements showed a clear tendency to manipulate available linguistic resources by increasing and even creating differences between neighbouring varieties in order to accentuate national identities (see Garvin 1964: 522, Hobsbawm 1990: 54, 101-111, Anderson 1991: 67-82). A language historian describing language variation in the context of nationalism needs to be careful not to make the mistake of accepting linguistic nationalism on its own terms, by uncritically taking the statements of nationalist leaders as reflecting facts about actual communication patterns; yet at the same time one has to acknowledge that nationalist leaders did not 'invent' traditions at whim. As Gal (1989: 349) has put it, it is vital to maintain an approach to the historical sources which 'navigate[s] between a radical distrust of language as a conspiratorial distortion and a relativist confidence in its neutrality'.

Afrikaner nationalism emerged in the mid 1870s in response to the challenge of British imperialism. The advent of Responsible Government in 1872 and the mineral revolution further intensified the struggle for political power and economic resources in the colony (Giliomee 1989). It was in this context that the first language society for 'Afrikaans', the Genootskap van Regte Afrikaners (GRA) was founded in 1875 in Paarl with the explicit aim of promoting the vernacular as a print-language (Schriftsprache, Viljoen 1891: 33f.), and as the national language of the Afrikaner nation.14 Soon after its foundation the GRA brought out its own periodical, Die

---

14 The original motivation which led to the foundation of the GRA is usually seen in Pannevis' suggestion to translate the Bible into what he had termed 'Afrikaans Hollands'. However, under the leadership of S.J. Du Toit a more nationalist orientation came to dominate, and without much discussion the GRA postponed the issue of the Bible translation after initial attempts to convince the BBBG (Briese en Buitenlandse Bybelgenootschap) failed. Du Plessis' (1986) interpretation of the GRA as a religiously
Afrikaanse Patriot, as well as the first grammar of Afrikaans (Eerste Beginseis van die Afrikaanse Taal, 1876). By 1895 over 70,000 copies of Afrikaans books and pamphlets, published by the society's own publishing house (D.F. du Toit & Co. founded in 1878) had been sold and the Patriot (now weekly) had a circulation of over 3000 copies (Du Toit 1909 [1880]: 78). The grammar (offered for a mere half shilling) also sold well, and by 1895 over 6000 copies had been sold (Du Toit 1909 [1880]: 173-176). The publication of a grammar was closely linked to the GRA's nationalist agenda, and aimed at establishing 'Afrikaans' as an autonomous language, which had its own rules and was not parasitic on the system of Dutch. Soon the type of language used by the GRA, which was highly uniform and characterised by almost complete regularisation of the verbal system, loss of gender, a pronoun system different from Dutch and certain innovative syntactic features, became known as Patriots Taal or Du Toit's Taal (the latter term reflects the fact that S.J. Du Toit had dominated the movement's codification efforts; see Dekker 1926: 77, Hollliday 1993: 15-17). Although no GRA-meetings were held after 1878 its leading members (especially S.J. Du Toit and C.P. Hoogenhout) continued with the publication of vernacular texts and codification efforts (including the second edition of the grammar 1897 and the publication of a dictionary 1902/1904).

From 1875 until the first language congress in Paarl in 1896 the GRA rejected (at least rhetorically) prescriptive standardisation and emphasised the primacy of linguistic usage; a position closely linked to S.J. Du Toit's language-programmatic writings in which he approaches language (following one prominent current of 19th century linguistic thought) as a natural organism (see Du Toit 1891: 16; see also AP 11/5/1877, p.73, AP 26/10/1877, p.169, 5/9/1895, p.2). This has led to an understanding of the codification work of the GRA as primarily descriptive (see for example Raids 1994 [1985]: 318). However, despite an overt laissez-faire rhetoric, poems sent to the Patriot for publication were corrected in accordance with the language use propagated by the GRA, and occasionally entire letters were rewritten (see, for example, AP 17/4/1890). Most contributors, however, followed the rules formulated in the 1876 grammar

inspired language movement whose activities were centred around the issue of the Bible translation is misleading in that it emphasises the initial impetus which was quickly (i.e. already in S.J. du Toit letters to the Zuid Afrikaan 1874) superseded by a nationalist discourse. Furthermore, Du Plessis who is concerned with separating language issues from non-linguistic agendas (such as spread of religion, promotion of a political goal), fails to see that 19th century romantic nationalism was per se language centred, i.e. language was not merely the instrument to spread the message, but it was part and parcel of this very message.

15 While Van Rensburg and Combrink (1984: 109) have claimed that the GRA model did not have much influence in the northern republics, letters sent to the Patriot suggest that the paper was read in the republics, an interpretation further supported by the fact that already in 1876 the Patriot had agents in the Free State and the Transvaal (see also Du Toit 1909 [1880]: 77). Note also the similarities between GRA-Afrikaans and previous vernacular writing (see § 8.1.3).
quite closely, limiting their linguistic experiments largely to questions of spelling. As with the vernacular writing tradition discussed in the previous section, to assess the extent to which 'Afrikaans' as promoted by the GRA showed signs of nationalist 'invention', one needs to consider comparative data on patterns of language use and variation.16

The GRA published vernacular reading material not only for adults but also for children: Hoogenhout's *Geskiedenis van Josef* was published as early as in 1873 (2nd edition in 1883), in 1878 followed the primer *Spel- en Leesboek ver Afrikaanse Kinders* (2nd edition in 1896), in 1880 the *Eerste Afrikaanse prinsjis boek ver soet kinders* (2nd edition in 1885) and the *Patriot* had a section titled *Ver Kinders*.17 The GRA also paid attention to what in today's terminology would be called 'acquisition planning' (Cooper 1989: 157-163) by publishing English-Afrikaans teaching material (*Samesprake in Afrikaans and Engels* (I-III) 1884-1890, second edition 1897; over 12.000 copies were sold before 1896, see *Di Tweede Afrikaanse Taalkongres* 1897: 4). 'Prestige planning' (on the term see Haarman 1990) was pursued not only through the publication of a grammar (thus bestowing on Afrikaans the prestige of a 'real' language), but also by appropriating arguments from popular linguistic science.

---

16 An alternative interpretation which, however, maintains the basic tenet of the creative nature of nationalist movements was put forward by Davids (1991, 1994), who has argued that 'Afrikaans' developed in its current form amongst the slaves and the Khoe, and was in the 1870s appropriated by the Afrikaners who then 'invented' a racially pure history for the language.

17 See for example AP 2/2/1877, 16/2/1877, 3/8/1877, 7/12/1877. This section was continued irregularly during the 1880s. See Bester (1989) for a general discussion of early Afrikaans children's literature.
Figure 3.2 Integration of Afrikaans in the Stammbaum of the Indoeuropean language family (Du Toit 1891: 3). Note that the assumed direct and close relationship between Afrikaans and the Germanic languages is visualised by the unchanged continuation of the Germanic language branch. The Dutch branch on the other hand, is much slighter and laterally shifted. The placement of English in between the Romance and the Germanic languages echoes the GRA's argument that English (and not Afrikaans) is a mengelmoes, i.e. a mixed language.

In the contemporary press the GRA members were often attacked as 'jong schoolmeesterjes en wynboertrieë' (Von Wielligh 1925b: 15), and on the whole the members of the GRA were prototypical supporters of 19th century nationalist movements, that is, the rising middle classes including the intellligentsia (Hobsbawm 1990: 116-122, Anderson 1991: 76, Smith 1991: 94-98). Independent farmers, ministers and teachers were prominent among the early members of the GRA, but there were also representatives of the newly emerging professional class (such J.M. Hoffman, a medical practitioner), as well as small entrepreneurs (such as D.G. Roussouw and P.S. Toerien, both shop keepers, I.A. Peroldt, a wagonmaker, G.W.Malherbe, a cart-builder, and J.J. Uys, a butcher; Cape of Good Hope Voter’s List 1903). All of them came from the Boland area (primarily Paarl, Wellington and Malmesbury) and their network ties were dense and multiplex. The majority of the society’s founding members were born in the 1840s and early 1850s, and this generational bias was maintained amongst those members who joined the society later (a list of members can be found following the Re’els and Bepalings, reprinted in Nienaber 1974: 55-56). It seems that although several farmers belonged to the movement, most of those who actively contributed to the Patriot were part of the
intelligentsia (see the remark in Du Toit 1909 [1880]: 77 on the need to find someone to write in the Patriot who knows about farming). In terms of their location in the social structure, the constituency of the GRA was thus remarkably similar to those who from the 1860s had supported and propagated the use of the vernacular for sociopolitical commentary.

After the institutional demise of the GRA from 1878, the nationalist agenda was taken over by the Afrikaner Bond which was a more politically centred nationalist organisation. The Bond was founded in 1879 by S.J. Du Toit to protect the economic and political interests of the petty bourgeoisie, and initially advocated a strong anti-English nationalism. The movement was 'hijacked' in the 1880s by wealthy commercial farmers to organise against the British imports of cheap wheat and brandy which threatened their profits, and under the leadership of Jan Hendrik Hofmeyr the Bond followed a more moderate policy collaborating with rather than challenging British dominance (Giliomee 1989: 66, Bickford-Smith 1995: 41). The change in constituency also affected the language-political orientation of the Bond, who now campaigned for the rights of Dutch in official functions while advocacy of 'Afrikaans' was limited to a minority of its members.

Standardisation efforts continued after the Anglo-Boer War (1899-1902), but were no longer in the hands of the GRA, and the geographical centre of the promotion of Afrikaans shifted from the Cape to the northern interior where Johannesburg and Pretoria established not only their position as economic but increasingly also as cultural centres. New language societies took over: the Afrikaanse Taalgeneeskaps (Transvaal) was founded in 1905, Onse Taal (OFS) and the Afrikaanse Taalvereniging (Cape) both in 1906. While much work of the GRA had centred on codification and the literature produced was folkloristic, the so-called 'second language movement' was rather literarily minded and the production of 'serious' literature became an important concern, leading to increasing popularity among the more traditional élite. The Zuidafrikaanse Akademie voor Taal, Letteren en Kunst was founded in 1909 with the explicit aim of promoting the Dutch languages in South Africa18, and following the introduction of Afrikaans in primary schools in 1914, the academy came to the fore as the central agency of language standardisation and elaboration. The time around 1917 has been described as a transition period (overgangstydperk) in which new standard norms were slowly implemented, but language use still exhibited considerable variation. Ultimately the elaboration of standard variants took place in the context of what has been termed Vernederlandsing, i.e. many lexical and syntactic variants which differed from continental Dutch were eliminated from the standard

---

18 The cultural and language-political tasks of the academy were specified as follows: 'De handhaving en bevordering van de hollandse Taal en Letteren en van de Zuid Afrikaanse Geschiedenis, Oudheidkunde en Kunst, en aan de ontwikkeling 'n wetenskappelike en artistieke leiding te geven. Onder het woord 'hollands' wordt verstaan de beide taalvormen gebruiklik in Zuidafrika.' (quoted in De Villiers 1934: 278).
language, leading to today's situation of 'a diglossic relation between the standard and spoken varieties of Afrikaans' (Ponelis 1993: 57; on the issue of Vernederlandsing see Valkhoff 1966: 239-240, Uys 1983).

3.3 Some Comments on Folk Taxonomies and Language Attitudes

Even a brief look at the extensive source material presented in Nienaber (1950) shows that visitors to the Cape as well as members of the Netherlandic speech community described the linguistic situation in terms of a rather static system of social and ethnic stratification, and commonly divided the speech community into discrete compartments paralleling ethnic and social groupings.

That the speakers' perceptions and opinions might contradict linguistic analysis (especially in continuum-type situations) and are dependent on variables such as age, social group, sex, etc. has been shown in empirical sociolinguistic research (Mattheier 1985, 1988). Language users rarely describe linguistic variation in continuous terms, but typically report the existence of a clear gap between two or three alternatives, often assigning everything which is short of the acrolectal norm to an amorphous 'vernacular' category. Linguists have, therefore, argued that what speakers think should not be confused with the system of varieties as actually in use (Bloomfield 1944, Grootaers 1959, Kloss 1978: 27, Klein & Dittmar 1979: 24, DeCamp 1971: 350). More recently, however, sociolinguists have begun to pay more serious attention to the speaker's overt descriptions of linguistic variation, as well as to folk-linguistic taxonomies. Such metalinguistically articulated (and socially mediated) beliefs about the structure of linguistic variation can be seen as belonging to Giddens' category of discursive knowledge, and thus enter into the (re-)production of sociolinguistic practices, including language maintenance, change and shift (see § 1.2.3).  

Common in the 19th century discourse at the Cape are binary classifications similar to

---

19 Bloomfield's (1964) comments on the categorisation of 'good' and 'bad' language use among the Menomini Indians shows that such binary categorisations are a universal phenomenon and do not necessarily depend on the availability of a standard variety against which derivations are measured. However, there is little doubt that in times of standardisation such categorising intensifies.

20 Nienaber (1950) who collected much source material reflecting the 19th century discourse on linguistic variation, makes in his own commentary the mistake of taking the folk taxonomies as direct evidence, and bases much of his sociohistorical analysis exclusively on this secondary data, leading to a rather compartmentalised view of linguistic variation. However, Nienaber also acknowledges that there existed a 'gradasie van afwyking tussen Afrikaans en Nederlands' (1950: 34).

21 See Woolard and Schieffelin (1994: 57-64), Gal (1995) as well as Preston's (1989) work on perceptual dialectology; a research programme for the study of folk linguistics was already outlined by Hoenigswald (1966), and it is central to the 'ethnography of speaking' (see Hymes 1974). On the use of 'people's accounts' in sociological research see Orbuch (1997).
Ross' (1954) classical distinction between U and non-U. Changuion (1844; 23) distinguished (excluding the metropolitan standard) educated from uneducated speech (the latter supposedly spoken primarily by the non-European population, but also common amongst artisans and farmers) based on the frequencies (sic!) of Africanderisms, i.e. linguistic forms which were believed to be typical for CDV, and unknown or rare in continental Dutch (see also Swaving 1830: 302f., Burchell 1953 [1822-1824]: 15, De Lima 1844: 9, Elffers 1903: 5). In the last quarter of the 19th century there existed a variety of terms used to describe basilectal CDV, all of which show that it was associated in the popular consciousness of the time with poverty and colour: Hotnotstaal, Griekwataal, Kombuistaal, Plattaal, Brabbeltaal and Bastaard Spraak (see Hofmeyer 1987: 97, Nienaber 1950: 24, 39-41). That the educated/upper classes did not speak 'Afrikaans' but rather an extraterritorial variety of Dutch was emphasised by commentators throughout the 19th century. M.L. Wessels described the situation in 1880 as follows:

'[A]s far as I am able to tell, the majority of the better class [...] speak High Dutch in a slightly modified form [...] in a form that differs from the purest Cape Dutch as day does from night. Their grammar may now and then be at fault, and the distinctions of grammar may be disregarded, but in the main their language is defective High Dutch. [...] That [...] the want of inflection is limited [...] to an indiscriminate use of the particle in one form to denote every case and gender, no one will deny; and it is only in the lower, I might say the lowest, the altogether uneducated classes, that the loss of inflection becomes generally perceptible in the verb.' (Cape Monthly Magazine 1880: 351, 45).

In an article written for the periodical De Zuid-Afrikaan (1/5/1890) Nicolaas Mansvelt distinguished Kombuis-Hollands from acrolectal Cape vernacular speech which he called Voorkamer- or Salon-Afrikaans. A more detailed description of the linguistic differences between educated (beskuit) and uneducated (plat) vernacular speech was presented in a letter published in the Zuid-Afrikaan (23/10/1890, also discussed in Nienaber 1950). The letter which claims to represent beskuit Afrikaans, shows a number of interesting linguistic features: maintenance of the preterite tense, absence of the 'double' negation; the adjective inflection still agrees with the Dutch system, the 1st person subject plural pronoun is wy; the 3rd person subject plural pronoun is sy, and plural verbs and infinitives are inflected. Uninflected plural verb forms and infinitives as well as the use of the pronoun ons in subject position are explicitly assigned to the system of plat Afrikaans.

While most commentators perceived only two forms of the vernacular (i.e. plat vs. beskuit), the GRA leader S.J. Du Toit distinguished three varieties: Heere-, 'Boere- and

22 Remarks on geographical differences are rare during the 19th and 20th centuries, but are discussed in some detail by Von Wielligh (1925a). Generally extraterritorial varieties seem to show less regional variation than their European relatives, which is possibly the result of strong dialect levelling tendencies in the early settlements.
Hottentots-Afrikaans (Zuid-Afrikaan 30/1/1875). This three-fold taxonomy was central to the language-political agenda of the GRA since it allowed to separate Boere-Afrikaans, which the GRA claimed was the language variety they promoted, not only from upper-class vernacular speech, but also from associations of colouredness and poverty (the three-fold scheme was, however, no invention of the GRA and had also occurred in earlier commentaries; see Nienaber 1950: 10-11, 25, also Elffers 1903: 5, Van Rijn 1914: 13). The Patriot was not only at pains to emphasise the difference between 'Afrikaans' and the vernacular of people of colour, but also ridiculed Heere-Afrikaans as an unsuccessful effort to speak or write Dutch, and rejected such linguistic varieties as 'unauthentic' (see, for example, AP 19/5/1882-3). While the variant nature of many varieties (maintaining, for example, certain inflections such -en for infinitive and plural marking, as well as the neuter for some nouns etc.) was thus evaluated negatively, the highly regularised and uniform language described in the first grammar was defined as 'authentic', as well as 'good' language. Thus, in the metalinguistic discourse of the GRA one witnesses the emergence of a puristic response to linguistic variation (a 'standard language ideology' in the Milroys' terminology), reflecting the need to establish clear, non-fuzzy boundaries between the linguistic norms of continental Dutch and the new national language 'Afrikaans'. By constructing 'Afrikaans' as a linguistic entity which could unambiguously be contrasted with continental Dutch, an important pre-requisite for the development of 'language loyalty' was achieved.

Perceptions of clear-cut boundaries and intra-language stability were however not shared by all of those who campaigned for vernacular elevation, as is evident from a letter (4/11/1894) by F.S. Malan to W. Viljoen.

Ek dink dat jou begrip van di Afrikaanse saak veul nouer en beperker is dan di van my. Di Afrikaanse taal wat jy voor werk is reeds gevorm in hoofsaak en het al meer van min syn bepaalde eigenaardighede ontvang. [...] Mar dan is daarn menigte ander Afrikaner wat meen dat di Afrikaans van vandag nog maar di eier in di nes is wat nog moet uitgebroei worde, 'n pas gebore kindje van wi jy nog ni kan se wat eigenlik di karaktertrekke van di man sal wees. [...] Jy het al di woord 'purist' met betrekking tot di jong taal, terwyl di anderkant meen dat alle woorde en

---

23 The need to disassociate 'Afrikaans' from the language variety used by people of colour is still evident in early 20th century texts (see for example Van Wieilagh 1925a: 96, see also Francken's historical play Susanna Reyniers (1908) in which the language of the slave character Plactus is described as a 'armoedig dialet' while that of the commander as 'beskaafd Afrikaans', for a discussion of the ideological implications of the play see Coetzee 1996).

24 On language loyalty as dependent on the availability of a clearly defined code and linguistic norms, see Weinreich (1968: 99-100), also Garvin (1964), Hill (1993); on the interconnectedness of linguistic purism, social movements and social change see Jernudd (1989), Annamalai (1989). Wexler (1971: 343) predicts linguistic purism in cases where the emerging standard is closely related to the old standard norm: 'In such a situation, we can expect widespread puristic efforts directed against any influences from the rival written norm'. For discussion of linguistic purism as 'a tool of dominance' see Hill (1985).
gebruikte ewe veul reg het om in di stryd van bestaan op te tre. (transcript of the letter in his diary 4/11/1894, Malan also refers to the 'onstandvastje karakter' of Afrikaans in other places, see for example diary entry 23/10/1894; a warning against a dogmatic approach to what constitutes Afrikaans was also issued by Malherbe 1917: 17-18).

The perception of GRA-Afrikaans as too extreme or exaggerated, and thus not representative of the variable nature of the actual spoken language, was not uncommon in the late 19th century (see, for example, the letter by Pietje Voorzichtig to the Zuid-Afrikaan 13/11/1875, see also the discussion in § 3.4).

An important aspect of linguistic folk taxonomies is that they constitute evaluative hierarchies and in that respect resemble (overt) language attitudes. It not surprising to find that in the context of growing racialisation of society (see Bickford-Smith 1995) a sociolinguistic category of 'Hottentot Afrikaans' became a focal part of the popular discourse on language use (especially since such a category was already part of the vernacular writing tradition). At the same time class demarcational needs are evident in the separation of educated vs. uneducated Afrikaans. The growing educational system furthermore helped to devalue popular modes of expression as 'patois', and at the same time supported the recognition of standard varieties as prestigious. In general, the attitudes expressed in contemporary discourse are similar to those surrounding many non-standard codes, varying between rejection as an example of degeneracy and acceptance as a symbol of solidarity.

3.4 A Diglossic Community?

The linguistic situation at the Cape during the 18th and 19th centuries is usually described as diglossic with CDV (or Afrikaans) as the L(ow) variety and Dutch as the H(igh) variety (spoken in formal contexts and written). In the sociolinguistic literature a distinction has been made between 'narrow' (or classical) diglossia and 'broad' diglossia. In speech communities characterised by 'narrow' diglossia two discrete varieties of one language exist in strict functional complementarity, the L variety is learned as the first language throughout the speech community, and the H variety is never used in informal settings (Ferguson 1959). In 'broad' diglossia, on the other hand, two or more styles, varieties or languages exist as stable elements in the speech community's communicative repertoire, each variety is allocated to different functions, yet there is no strict complementarity, and H can occur in informal conversations; the acquisition of L as

a first language is not at issue (Fishman 1967/1980, Fasold 1984: 52-54, Myers-Scotton 1986). 'Broad' diglossia thus not only includes cases where H and L are genetically unrelated languages, but is also used to classify situations where the speech community is disjunctive, that is, not all individuals in the community show diglossic behaviour, however, the overall societal distribution appears to be diglossic (see Fishman 1980: 7-8). While the extension of the notion of 'diglossia' to scenarios where the languages in question are unrelated is generally accepted, extending the term to disjunctive speech communities is seen as problematic (Timm 1981). In order to keep the different scenarios conceptually separate, Britto (1986: 35-40, 287, see also Schönfeld 1985: 211 for a similar classification) suggested the following classification of diglossic situations:

(i) **use-oriented diglossia**: classical diglossia as described by Ferguson, that is, H is superposed for the whole community (i.e. everyone learns H as an L2), the use of H depends on domain specialisation and is thus use oriented; no section of the community uses H for ordinary conversation.

(ii) **dialectal or user-oriented diglossia**: H is not a superposed variety for all members of the speech community (i.e. certain groups acquire H as their L1 and use it in informal conversation). As a result use of H/L depends not only on domain, but also on social characteristics such as ethnicity, religion or class (this type of diglossia roughly corresponds to Ferguson's category 'standard-with-dialects').

(iii) **pseudo diglossia**: the two varieties or languages are used by separate speech communities within a given geographical or political organisation; there is, however, no within-group diglossia.26

To view the society at the Cape as diglossic has important consequences for the interpretation of the source material, since in cases of use-oriented diglossia the written language typically belongs to the H domain, and thus offers the language historian little evidence for the reconstruction of patterns of variation in the spoken language. The situation is different in the case of dialectal diglossia, where two (or more) clearly differentiated codes exist at the level of the speech community, but not necessarily at the level of the individual.

During the late 19th and early 20th century visitors to the Cape as well as members of the speech community frequently remarked on the existence of a gap between the casual spoken and

---

26 Much of the discussion surrounding the question of 'what' qualifies as diglossia (as well as the general fuzziness of the term sharply attacked by Dittmar 1976: 178) is related to the fact that Ferguson's original paper was rather tentative and exploratory, as explicitly acknowledged in his closing remarks: 'Perhaps the collection of data and more profound study will drastically modify the impressionistic remarks of this paper, but if this is so the paper will have had the virtue of stimulating investigation and thought.' (1959: 340).
the more formal spoken and especially written registers (see, for example, Brill 1909 [1875], Colquhoun 1906: 139-142, De Vooys 1913: 179, Langenhoven 1914: 38-34, 52-53). However, in interpreting such contemporary descriptions one has to take account of the fact that from the mid 19th century references to a gap between spoken and written language had become a commonplace in academic as well as popular linguistic discourse, which began to emphasise the primacy of the spoken language, while the philologists' preoccupation with written sources (now branded as 'artificial') came under attack (Scaglione 1984: 41, Steinberg 1987: 201-202, Joseph 1989: 252-253). In the Netherlands as well as elsewhere in Europe grammarians increasingly focused on the differences between spoken and written language, and soon the oral-written gap was identified as a 'social problem' believed to limit children's success in school, and keeping adult literacy levels low (Moledijk 1992: 170, 173). At the Cape references to a gap between spoken and written/formal language were particularly common in the writings of those supporting the nationalist agenda of the GRA (see, for example, AP 31/8/1877, 22/5/1890), and by emphasising the existence of extensive differences between the 'vernacular' and formal Dutch (which were said to limit access to reading material and education) they succeeded in legitimising the project of vernacular elevation (on strategies for legitimising claims and action, see Blumer 1971: 303-304, Pomerantz 1986). Other commentators, however, insisted that the problems people encountered when reading Dutch were not so much a question of diverging grammars but rather of style, i.e. what was needed was not vernacular elevation but the use of a more simple style, a kind of 'plain' Dutch (see, for example, letter by 'conservatief' to the Zuid Afrikaan 24/2/1875, see also the text quoted in De Villiers 1934: 177-176).

While to emphasise the gap between spoken and written (formal) varieties was to some extent the result of the popular as well as the scientific Zeitgeist, detailed remarks on the functional distribution suggest that the notion of diglossia might be applicable to the linguistic situation at the Cape. Thus, Viljoen's (1895: 24) description of the use of CDV is typical of an L variety.

---

27 See, for example, Roorda (1856: 97ff.), Verdam (1890: 57), Malssen (1900: 191-213), Te Winkel (1904: 279); for an overview see Noordegraaf (1985: 339-343, 412). It is a truism of sociological research that 'social problems' (i.e. problems believed to be harmful to society as a whole) do not necessarily correspond to 'objectively' harmful realities. Relatively benign conditions are frequently catapulted into the floodlight of social and political attention, while extremely harmful conditions are often ignored (Goode & Nachman 1994: 94). Thus, Moledijk remarked that in the 19th century when academics and educationalists argued that the Dutch spelling rules obstructed the acquisition of literacy, illiteracy was in reality at an all time low (Moledijk 1992: 170).

28 Interesting is also that many locally produced primers (such as De Villier's Fondemenssteenen 1884-1894) explicitly comment on diverging pronunciation features (such as /sk/ for /sx/ and apocope of /t/) and note some lexical peculiarities, but rarely remark on divergent grammatical and syntactic features in the children's vernacular.

Dutch, on the other hand, is 'Kirchen-und Kanzelsprache', used in prayer, writing and formal situations (Viljoen 1896: 21-22, Te Winkel 1897: 10, Hoogenhout 1904: 913: also Nienaber 1950: 94-95). It is noticeable, however, that while travellers had referred to the existence of a Cape Dutch dialect from the mid-18th century (Mentzel, Sparrman, Burchell, Swaving, Lichtenstein, Teenstra), the conceptual distinction of a clear H/L dichotomy surfaces in the secondary sources only from about 1850 (and thus developed in parallel to the development of the vernacular writing tradition, and the emergence of the label 'Afrikaans'). In this context, the case of Romance can be instructive. Traditionally language historians have assumed the existence of a Latin-(proto) Romance diglossia for the Early Middle Ages, with the two varieties conceptually and structurally distinct. This model has been questioned more recently by Wright (1982, 1991) who argued that the conceptual distinction between Latin and Romance (vernacular) was 'created' by the Carolingian scholars when they developed a new system of distinctively non-Latin spellings to represent certain varieties, which were subsequently seen as distinct L-varieties. Thus, while during the eighth and ninth centuries there existed a Latin/Romance variation continuum, this was transformed under the influence of the Carolingian linguistic reform into an opposition of languages (Latin (H)- Romance (L)) in the Later Middle Ages. In his discussion Wright (1991: 107) makes an important comment with respect to diglossia, namely, that diglossia does not evolve naturally, but necessitates human intervention:

[S]ystems of diglossia need not only be intentionally set up, but also to be continually reinforced subsequently - mainly by teaching the high variety in the education system - in order to arise at all, and do not arise naturally otherwise. What exists otherwise, and do indeed evolve unplanned, in a single wide speech community, are complex patterns of sociolinguistic variation.

This also implies that within a single speech community different groups of speakers will be affected to differing degrees by such 'diglossic engineering', i.e. it affects first and foremost those with regular access to education, and these were few in the colony (see § 4.1 on the education system in the 19th century). Since our secondary sources are biased (that is, they reflect primarily the perceptions of the educated classes, who expressed their opinions in newspapers and books) careful interpretation of this evidence is necessary. Furthermore, the neat picture of functional specialisation described by Viljoen, becomes rather more ambiguous when one considers that not all members of the speech community described the informal, spoken language as a uniform category. Several commentators emphasised that the spoken vernacular comprised different varieties some of which were reportedly relatively close to the metropolitan standard,
and were used by the 'upper classes' in casual speech. Given such references to various intermediate forms of the spoken language it is debatable whether the situation at the Cape can or should be described as diglossic, or whether these observations indicate the possibility of a dialect continuum.

Although Ferguson's definition of diglossia has emphasised the bi-systemic nature of the situation (H and L as two non-overlapping sets of linguistic forms), he acknowledged the possibility that intermediate varieties might develop (1959: 332). However, according to Britto's (1986: 15ff.) interpretation of Ferguson's original position these intermediate forms are usually clearly situated within the 'linguistic duality' or bi-systemicity which is the defining characteristic of diglossia, i.e. these varieties are either part of H or L. Only under the following conditions should they be classified as an independent diasytem:

[I]t seems justifiable to recognize more than two diasystems in diglossia, provided there is Optimal difference [i.e. definitely more distance than between styles, but less than between separate languages] among the diasystems not only in linguistic structure but also in function and possibly other parameters such as acquisition, prestige, literary heritage, etc. The condition of Optimal distance is necessary to avoid the recognition of every variety or code as a diasytem and every subtle difference in function, acquisition, prestige, etc. as significant, which would make the theory of diglossia as Ferguson envisioned it superfluous (Britto: 1986: 18, on Britto's criterion of Optimal distance see § 8.1.4, on the need to relax the criterion of 'binarity' see also Fasold 1984: 53).

Similarly Fishman (1980: 4) remarked that the structural gap between the different systems or linguistic codes needs to be such 'that without schooling the written/formal-spoken cannot be understood (otherwise every dialect/standard situation in the world would qualify within this rubric)' [my emphasis], and Wexler (1971: 336) emphasised that 'diglossia is not meant to refer to any condition of multiple norms, but specifically to that condition where there is a broad structural gap' correlating with patterns of functional complementarity (also Keller 1982: 73, Winford 1984: 349). While the requirements for 'true' diglossia are thus relatively strict (structural discontinuity, strict functional complementarity and L variety is always acquired as L1), Fishman has emphasised that diglossia (like many other social constructs) is a continuous variable, a matter of degree rather than type (1980: 6), a concern also reflected in Schiffman's

---

29 See § 3.3; see, for example, also the evidence given by Prof. Van der Turk to the education commission (First Report 1891: 102), De Wet (1876 [quoted in Du Toit 1909/1880]: 102-103), Besselaar (1914: 193-194, 219-220), Van Rijn (1914: 13), as well as the testimony by J.J. Willemsen (quoted in Nienaber 1950: 34-35), and the letter by W. to De Goede Hoop (November, 1906). The Cape-born student C.H. Persoon who went to study in the Netherlands, reported (in 1775) in a letters to is parents that frequent comments were made about his Caaps krom spreken (see Raidt 1983: 40). This can be interpreted as indicating the existence of a marked extraterritorial variety, while the linguistic differences (although sufficient for ridicule) seem to have been small enough not to lead to communication problems.
(1996: 27) phrase 'diglossic tendencies'. However, while it is important to realise that diglossia is not a $[\pm]$ feature, there is the continuing danger of diverging too far from Ferguson's original discussion, and being left with a definition of diglossia which is applicable to any speech community. The question of whether the intermediate forms described in the contemporary sources warrant the classification as an independent system M, whether they were part of H, or whether there existed a finely graded linguistic continuum between Dutch and more basilectal varieties of CDV will be discussed in more detail in chapter 8.

Varieties close to the Dutch norm appear to have been prominent enough to make the idea of promoting a simplified, local form of Dutch as the future standard language a viable project, which was supported not only by the church but also by several language professionals (such as Mansvelt, Brill and Viljoen). The foundation of a society for the promotion of Dutch was first suggested by Jan Hendrik Hofmeyer in 1877, but nothing came of the proposed organisation for Dutch until the 1880s when the aggressive nature of British imperialism ensured not only the political organisation of the Afrikaner population in the Bond but also the cultural-linguistic mobilisation of the Dutch-speaking section of the community. Dutch had been excluded from official functions and the state-sponsored school system since 1822, when a government proclamation granted official status to English only (see § 8.2.1). In 1881 a petition signed by over 6000 individuals was handed to the government, and in 1882 Dutch was re-instated in Parliament and the school system, in 1884 in the Courts of Justice, and in 1888 Dutch was made an obligatory subject for civil servants (De Villiers 1934: 109-120).

The language variety promoted from the mid 1890s by the supporters of Dutch (who founded the Zuid-Afrikaansche Taalbond in 1890) was not continental Dutch but a simplified Cape Dutch (eenvoudig Kaapsch Hollandsch) which was said to differ from 'Afrikaans' as promoted by the GRA. Eenvoudig Kaapsch Hollandsch was supposedly characterised by indigenous vocabulary items, loss of case inflection in the article and increasing loss of the neuter; the conjugation of verbs followed largely the paradigm of weak verbs and a partially different pronominal system was used (De Vos 1891: 6, ZAT 1892: 162-165, Engelenburg 1897: 360, also Mansvelt in The Cape Times 4/11/1890). The first congress for the simplification of Dutch took

---

30 It has been suggested that the relatively strong position of varieties close to Dutch in the 19th century was stimulated through a 'renaisssance' of Dutch resulting from increasing Dutch immigration after 1850 (Ponelis 1993: 47). However, although Dutch immigrants played an important role in 19th century language political discourse (Pannevis, Hoogenhout, Mansvelt), their numbers were on the whole marginal. According to the census (1891: 80) only 0.23% of persons born outside the colony came from the Netherlands. The total percentage of European born persons of the White population was 13.2%, the vast majority of which came from the British Isles. Raait (1991: 241) attributes the 'renaissance' of Dutch not so much to immigration patterns, but to the fact that with increasing attention to educational questions from the 1830s, the question of language standards (in this case the standard of continental Dutch) became an important issue (see Changurion 1844).
place in 1897 in Cape Town. Participants came not only from the Cape itself but also from the OFS and the Transvaal Republic. The discussion of simplification was influenced significantly by the language reform movement in the Netherlands where Kollewijn and his supporters had put the question of orthographical and grammatical simplification on the agenda (see Kollewijn 1903). The newspaper *Ons Land* (founded by J.J.Hofmeyer in 1892, united with the *Zuid-Afrikaan* 1893) as well as several South African writers advocated the use of a simplified form of Dutch (De Villiers 1934: 161). Frequently cited in this context is the position of the writer Melt Brink (1842-1915), an autodidact with little formal schooling, who wrote some of his work in what he described as a *tussentaalvorm*, located somewhere between Dutch and 'Afrikaans' (see also Te Winkel 1896: 432, Meyer-Benfey 1904: 232, D'Arbez (J. Van Oortt) in *Die Brandwag* 1/10/1914).

Ik heb mij, wat de taal betreft, zooveel mij zulks mogelijk was, aan de middenweg tussen goed Hollands en Patriots gehoude. De eerste omdat ik denk dat deze voor velen gemakkelijker zal lezen, en ten tweede omdat ik het Patriots te overbrengen acht, *wij spreken zo niet onder ons*. Ik heb het Kaaps-Hollands, zoals wij het gewoonlijk spreken, gevolgd' (quoted in Kannemeyer 1984: 71, my emphasis).\(^{31}\)

A central problem of the *Taalbond* and its efforts to establish a simplified South African Dutch as the language of education and administration, was that codification was delayed as its supporters did not want to act without backing from the *Vereenwoordigers* in the Netherlands (*Verslag van Het Eerste Congres* 1897: 11, 17; Engelenburg 1897: 362f., De Villiers 1934: 171, see also the commentary in the *AP* 26/12/1895, 6/2/1896). When the simplification proposals were finally accepted they were relatively modest (focusing largely only on issues of spelling), and normative publications in the newly established standard did not appear before 1907. As a result of these delays, the variety propagated by the *Taalbond* remained diffuse, leading to considerable insecurity regarding accepted usage; a situation not acceptable in a time when the challenge was to gain control of the H domain (see § 3.5). The lack of a clear linguistic type was exploited by those promoting Afrikaans who dubbed Cape Dutch a language with 'gen kop of stert' (*AP* 24/12/1891-2, also *AP* 9/10/1890-2, also *Cape Times* 17/11/1890). The general insecurity about

---

\(^{31}\) Similar intermediate forms appeared in the periodical press. Interesting in this respect is not only *Ons Land*, but also *The Friend of the Free State* (see Raidt 1994 [1922]) and the *APO* paper (see Adhikari 1996) which published a popular column (*Straatpraatjes*) written in 'Afrikaans' (including many linguistic features typical for Cape vernacular speech) as well as articles in what Ponelis (1996: 134) has called 'approximate Afrikaans', i.e. an intermediate form located somewhere on a continuum between Dutch and Afrikaans (see § 8.1.3). In the Muslim CDV literature the use of intermediate forms (termed *gemixte taal* by Davids 1991) was common in publications written in Roman script, while Arabic Afrikaans texts (i.e. texts written in the Arabic script) conformed from the 1860s rather closely to the model of 'Afrikaans' as known from the vernacular writing tradition (Ponelis 1981, Davids 1991: 96ff).
the linguistic status of Cape Dutch is aptly expressed in a poem published in the ZAT in February 1890, which shows that Cape Dutch could be identified only by stating what it is not (i.e. neither Dutch nor Afrikaans), but that no positive definition was available.

Ek zou nou graag wille weet,
Wat ver 'n taal ons hier praat,
Want Afrikaans det is 't niet
Wat ek daarvan verstaat.

Ook is dit nie die paterjot
Ni Boesman, ni Koraan
Ni Portegeis, ni Makatees,
Ni Kaffir, ni Javaan.

Di mense di det hollans noem,
Is glad en al verkeerd,
Geen hollander die praat en skrijf
As ons det word geleerd. (ZAT, February 1890: 17)

In sum, the secondary responses suggest that the Netherlandic speech community at the Cape showed a number of diglossic tendencies. There is little doubt that continental standard Dutch served as a kind of 'dummy H', i.e. it was seen as prestigious, learnt through secondary exposure (i.e. reading and formal schooling), but was rarely used by Cape-born speakers except in official documents (on the concept of 'Dummy High' see Platt 1977: 373-374). The growing prominence of nationalism in the 19th century was furthermore at odds with the maintenance of an exogenous H variety. There is some evidence that from the second half of the 19th century the situation at the Cape was changing rather rapidly, and that the knowledge and use of the norms of continental Dutch was fading; a development which was possibly accelerated by the existence of an English medium public school system from the 1830s (Mansveld 1884: 180, see also Wessels in Cape Monthly Magazine 1880: 9, Viljoen 1891: 4, see also § 8.2.1).

Some forms of CDV (beskaffer) were prestigious enough to be used in (spoken) H domains such as sermons and political speeches (see Verslag van het Eerste Congres 1897: 6, see also Te Winkel 1897: 10, Elffers 1903: 532, Colquhoun 1906: 141; similarly 'educated spoken Arabic' is used in sermons, speeches and letters, El-Hassan 1977: 113-114, 116), thus constituting alternative

32 Elffers (ibid.) described such acrolectal varieties rather impressionistically as follows: 'a language lacking the grammatical niceties of the Dutch of Holland and the shades of meaning which necessarily adorn a tongue of which the learned make use, besides discarding much of the idiom of the North - but none the less expressive'. He distinguished these from the 'real patois': 'low and undeveloped, dependent on circumstances and locality, easily influenced, and becoming more and more Anglicised'.
forms of H.\textsuperscript{33} Basilectal CDV (\textit{plat}), on the other hand, is a typical example of a stigmatised L variety, and terms such as \textit{Kombuistaal} unambiguously refer to its identity in the popular consciousness as well as its social domain.

Our secondary sources also report that varieties relatively close to Dutch were used by urban, educated members of the speech community in informal and casual interaction, while their use of basilectal CDV appears to have been limited to a rather restrictive set of domains (generally socially non-symmetrical interactions), which supports an interpretation of the situation at the Cape as an example of user-oriented or dialectal diglossia, i.e. H (and its varieties) were not superposed for the entire speech community. Furthermore, since the systemic status of the described intermediate varieties is ambiguous (i.e. they appear to be neither H nor L), the possibility of a variation continuum cannot be dismissed. Those commentators who emphasised the existence of intermediate forms often indicated that the nature of CDV speech was more fluid and continuous than reflected by the popular two- and three-fold taxonomies which dominated much of the metalinguistic discourse. A more continuum-type situation is suggested, for example, by Colquhoun's statement that 'the most puzzling feature of this patois is found in the shades or types of vernacular which are encountered.' (1906: 139; Besselaar 1914: 205, 211; see also Elffers' 1903: 6-7 discussion of the 'Shades of language and where to expect them').

The characteristics of user-oriented or dialectal diglossia (or according to Ferguson a 'standard-with-dialects' situation) were summarised eloquently by Paul in his \textit{Prinzipien der Sprachgeschichte} (1920: 411-412, see Macha 1992 for an empirical study of the complex patterns of language use and variation in dialect-standard scenarios):

\begin{quote}
In jedem Gebiete, für welches eine gemeinsprachliche Norm besteht, zeigen sich die Sprachen der einzelnen Individuen als sehr mannigfache Abstufungen. Zwischen denen, welche der Norm so nahe als möglich kommen, und denen, welche die verschiedensten Mundarten am wenigsten von der Norm inficiert darstellen, gibt es viele Vermittlungen. Dabei verwenden die meisten Individuen zweii, mitunter sogar noch mehr Sprachen, von denen die eine der Norm, die andere der Mundart näher steht. [...] Es kommen natürlich auch Individuen vor, die sich nur einer Sprache bedienen, weder solche, die in ihrer natürlichen Sprache der Norm schon so nahe kommen oder zu kommen glauben, dass sie es nicht mehr für nötig halten sich derselben durch künstliche Bemühungen noch weiter zu nähern, andererseits solche, die von den Bedürfnissen noch unberührt sind, die zur Schöpfung und Anwendung der Gemeinsprache geführt haben.
\end{quote}

Irrespective, however, of whether one assumes (classical) diglossia or standard-with-

\textsuperscript{33} Note in this context the following statement by Malherbe (1917: 18) which illustrates not only the growing vernacular purism discussed in § 3.3 but also emphasises the continuing influence of Dutch norms on the use of Afrikaans in H domains: 'Ook van die kant van Nederlands dreig 'n gevaar. Ons weet hoe moeilik dit vir 'n openbare spreker was wat gewoon geraak het aan \textit{sy soort Nederlands} om \textit{'n Afrikaanse aanspraak te lewe. Hy verval onbewus in Nederlandske vorme. Dit geld ook van schrywe} [my emphasis].
dialects as the appropriate scenario for the Cape, the labels H and L are useful shorthands (for functional domains as well as styles or varieties within an individual’s repertoire). and just as the creolist’s terminology ofacro-, meso- and basilect can be useful in non-creole settings, the use of the labels H and L can be useful outside strictly diglossic situations.

3.5 The Linguistic 'Marketplace' and its Entrepreneurs

During the 1980s and 1990s sociolinguists began to conceptualise linguistic variation in terms of the linguistic 'marketplace' (see for example Sankoff & Laberge 1978, Dittmar, Schlobinski & Wachs 1988, Gal 1989, Coulmas 1992: 58ff.). Linguistic markets resemble economic markets in that their structure is under no one's deliberate control, but is the cumulative outcome of the activities of its participants (i.e. they are the result of 'invisible hand processes'. see §1.2.3). Deliberate intervention is possible, yet the results of such interventions are predictable only within limits (on the structure of markets in general see Schelling 1972: 17). The metaphor of the linguistic 'marketplace' is useful as it allows one to link linguistic practices to issues of (social and political) power and authority. That is, not all forms of language have the same 'value' in the market since their value depends on the power and authority of the groups they index (Bourdieu 1991: 67). The unification of the market is, however, never complete and on alternative markets different hierarchies exist (Woolard 1985: 743-746, Bourdieu 1991: 71). In this sense, linguistic variants or varieties which are usually described as commanding 'covert' prestige' can be conceptualised as dominating alternative markets, that is markets in which the law of 'price formation' is linked to a solidarity rather than a status index.

Using the market metaphor one might call those who try to manipulate the structural organisation of the market 'language entrepreneurs', a term introduced by Rubin (1977; Weinstein 1979, 1982 suggested the more politically inspired term 'language strategists'). While at any time the linguistic repertoire of a speech community comprises a variety of ways of speaking which are not necessarily sorted into a clear-cut system of codes and language-related identities, 'language entrepreneurs' attempt to shape the heteroglossia they encounter into 'voices' that constitute community identity (Bakhtin 1981: 272, 293-294,356-357, also Hill 1993). The members of both the GRA and the Taalbond can be described as exemplary of such 'language entrepreneurs'. Both language societies tried to change through collective action the structure of the formal linguistic market (H domain) by challenging the historical legitimacy of continental Dutch. and by trying to establish the legitimacy of a new national standard language (Cape Dutch and Afrikaans respectively).

Linguistic nationalism, despite its rhetorical attention to 'the people', has always targeted

---

34 Unlike Wright's (1994) 'trendsetters', 'language entrepreneurs' are on the whole conscious of their actions and reflect explicitly on their goals and methods.
first and foremost the formal linguistic market, aiming to win official recognition (and thus legitimacy and 'overt prestige') for the new 'national language' (see Hobsbawm 1990: 110), and it is in this context that Gellner (1983: 35-38) formulated his hypothesis that 19th century announced what he has called 'the age of a universal high culture'.

[N]ationalism is, essentially, the general imposition of a high culture on society, where previously low cultures had taken up the lives of the majority, and in some cases of the totality of the population. It means the generalized diffusion of a school-mediated, academy-supervised idiom. codified for the requirements of reasonably precise bureaucratic and technological communication. It is the establishment of an anonymous, impersonal society, with mutually substitutable atomized individuals, held together above all by a shared culture of this kind, in place of a previous complex structure of local groups, sustained by folk cultures reproduced locally and idiosyncratically by the micro-group themselves. [...] If the nationalism prospers it eliminates the alien high culture, but it does not replace it by the old local low culture; it revives, or invents, a local high (literate, specialist-transmitted) culture of its own, though admittedly one which will have some links with the earlier local folk styles and dialects. But it was the great ladies at the Budapest Opera who really went to town in peasant dresses, or dresses claimed to be such. (Gellner 1983: 57)

From this perspective the conflict between the GRA and the Taalbond was a struggle for cultural hegemony (in Gramsci's terminology) or symbolic power (in Bourdieu's terminology). a struggle for the right to determine what constituted the legitimate and authoritative language of the 'Afrikaner nation'. At the same time the (often highly variable) practices of 'the people' were marginalised and devalued as irregular, unauthentic, and undesirable. Thus, the term 'vernacular elevation' does not refer to the acceptance of sociolinguistic heteroglossia as 'the standard', but typically describes a process of H-replacement, i.e. it is only the identity of H which changes, not its nature. That the new H, characterised by the uniformity and normativity typical for standard languages, is created according to the model of the old H, was also emphasised by Joseph (1987: 52, 106) in his discussion of the 'avantgarde of acculturation':

[T]hose who would undertake the standardization of L seem to betray L culture [...]. They 'sell out' L for a position within H culture that is certain to elevate their own personal prestige. [...] Thus, although the work of the L cultural avantgarde in adapting the synecdochic dialect for standard language functions is typically motivated by a desire to promote the instruction and acculturation of the L community at large, the very adoption they perform alters the L language - by standardizing it- such that it is not longer the language of the mass of the L community.' [my emphasis].

Holliday (1993: 26-27) has interpreted the competition between the GRA and the Taalbond in terms of a conflict between traditional élite (Taalbond, supported primarily by the 'grand bourgeoisie', the church and 'Gelehrten'; Viljoen 1896: 23) and counter-élite (GRA, supported primarily by the emerging professional class and the petty bourgeoisie). That counter-
élites often represent new social and cultural interests, which react against the established norms of tradition has been shown repeatedly in sociohistorical research (Joseph 1987: 51-56. Bottomore 1993[1964]: 8). However, despite general differences in constituency and goals, personal continuities existed between the GRA and the Taalbond. Supporters of Dutch were present at the first language congress for Afrikaans, and members of the GRA actively supported efforts to re-establish Dutch in official functions in the colony, e.g. C.P. Hoogenhout, one of the founder members of the GRA, was vice-chair of the Taalbond in Wellington, other GRA members involved in the work of the Taalbond were D.F. Du Toit and S.J. du Toit (De Villiers 1934: 134, 151). Furthermore, the GRA never positioned itself openly in opposition to the efforts to reinstate Dutch (or varieties close to Dutch), but supported (parallel to the political unification in the Bond) a broad opposition to English (AP 11/5/1877, 5/3/1891 'Net maar teeno'er Engels moet ons eensgesind bly en ni onder mekaar stry', also AP 1/7/1881, 6/5/1887).35 Such seemingly contradictory linguistic loyalties have been described in detail by Fishman (1987) in his biography of Nathan Birnbaum (who supported first Hebrew, then Yiddish and finally no particular language as a marker of Jewish identity), and based on his analysis Fishman urged researchers 'to question the usual functional analysis of elitist language ideology in terms of self-serving or class-related bias' (1987: 129). Thus, while a narrow reading of, for example, Gramsci or Bourdieu could be seen as suggesting a direct and unidimensional relation between code manipulations and changes or struggles in power or class relations, detailed sociohistorical research reveals a more complex and variable picture.

In sum, the second half of the 19th century saw what historians call the questione della lingua (i.e. a complex combination of issues about language and power) rise to prominence in the popular and political discourse. Questions of code identification and differentiation - i.e. what counts as a language and what are its boundaries - featured prominently in the metalinguistic discourse, and were manipulated by individuals and groups. Contemporary folk taxonomies suggest that the linguistic situation was often perceived as relatively focused, although there are indications that the actual patterns of variation were more fluid. For most language activists it was a question of either 'Dutch' or 'Afrikaans', while the ambiguity of possible in-betweens was rejected, ridiculed and branded as non-authentic. The mental make-up of the time resembles what Zerubavel (1991) has termed the 'rigid mind', a mind which shuns the greys, the twilight, and perceives the world (in Aristotelian fashion) as a composite of clear-cut entities.

The most distinctive characteristic of the rigid mind is its unyielding commitment to the mutual

35 At the same time, however, subtle criticisms were levelled at the supporters of the Dutch movement, such as when it was claimed that those who supported Dutch (rather than Afrikaans) actually spoke English at home (AP 8/5/1890-2).
The foremost logical prerequisite of rigid classification is that a mental item belongs to no more than one category. (Zerubavel 1991: 34)

Nationalist movements per se are exemplifications of the rigid mind, as the setting of boundaries, the unambiguous definition of the 'us' and 'they' is vital to their cultural and political agenda. The symbolic association between language and group depends on a set of ideas whose most essential condition is the belief that both groups of people and forms of speech exist as distinct entities; yet the process of definition, the marking of the boundaries, is generally obscured in the nationalist discourse which (as argued in § 3.2 and § 3.3) propagates the 'naturalness' of the distinctions made. The sociolinguist investigating linguistic variation and change under such historical conditions needs to be aware of the fact that conscious manipulation of linguistic resources is pervasive during such times, and that folk taxonomies, rather than describing the realities of language use, must be understood as part of a discourse reflecting the struggle for domination. As Foucault (1984: 110) reminds us: 'discourse is not simply that which translates struggles or systems of domination, but is the thing for which and by which there is struggle, discourse is the power which is to be seized'.
PART II

DATA COLLECTION AND STATISTICAL METHODOLOGY
THE SURVEY: SOCIOLINGUISTIC VARIATION IN THE CAPE COLONY (1880-1922)

Explicit theorising is an attempt to be intellectually honest. It is a way of penetrating the structure of language and the habits of thought - the image-reservoirs and shimmer of words - which drown our understanding.

R.S. Neale Class and Class Consciousness in Early Nineteenth Century England (1968)

4.1 Constructing the Sociohistorical Sample

4.1.1 The Netherlandic Speech Community

The importance of representative sampling for variation analysis has been emphasised by Labovian sociolinguistics. The use of random samples is rare in sociolinguistic research, and most studies have been based on stratified (or judgement) samples in which the researcher identifies in advance the social dimensions that are believed to be important for the study (such as age, sex, region etc.), and selects an appropriate number of speakers from each group (e.g. an equal number of men and women, of young, middle-aged and old people, and so forth) to ensure representativity.

The sociolinguistic sample is meant to be representative of that which is the starting point of the analysis: the 'speech community' - a category which despite its conceptual fuzziness is central to the sociolinguistic enterprise. While some have approached the speech community from a strictly linguistic point of view (Lyons 1970: 326, 'all people who use a given language (or dialect)'), others have emphasised its regional and cultural dimensions (Frings 1957), its interaction patterns and shared knowledge (see Gumperz 1962/1997), or its linguistic and sociopsychological commonalities (Labov 1991 [1972]: 120-21). The Labovian definition of the 'speech community' as a group of speakers who share the same evaluation of the linguistic variants (homogeneity of interpretation) which differentiate the speakers socially (heterogeneity of production; Labov 1982: 18), has been central to variation studies. On several counts, however, Labov's model has attracted criticism. While the model has been useful for the sociolinguistic analysis of monolingual American urban settings, it has been criticised for failing to address other common types of language settings, such as multilingual societies and societies influenced by large-scale migration (Romaine 1982b).

Labov explicitly excluded non-native speakers of the urban dialect from both his NYC study and his Philadelphia research (Labov 1980: 373).\(^1\) The legitimacy of this procedure was

---

\(^1\) The cut off-point for 'nativeness' is usually around the age of eight as children who came to the area after the age of eight rarely show the defining characteristics of the speech community (Labov 1980: 373).
questioned by Horvath (1985) who included different ethnic groups into her sample of the Sydney speech community. Her results, however, give empirical support to Labov's underlying hypothesis of the marginality of non-native speakers. That is, first generation immigrants to Sydney were shown to form a peripheral speech community, characterised by the use of variants which Horvath classified as 'Ethnic Broad', and which were absent from the linguistic behaviour of members of the core speech community. Second generation immigrants, however, were found to be full members of the core speech community.

Bilingual communities, such as the Gaelic and English speaking community in East Sutherland described by Dorian, pose the most significant challenge to the Labovian model of the speech community. Dorian has emphasised that in East Sutherland, marginal individuals such as 'low proficiency 'semi-speakers' and near-passive bilinguals in Gaelic and English' (Dorian 1982: 26) who lack the linguistic knowledge of important linguistic variants, and who do not participate in the patterns of social evaluation shared by the core speakers, are nevertheless seen by others and themselves as integral members of the Gaelic speech community. Differing degrees of proficiency also characterise the French-English bilingual speech community in Canada. Mougeon & Beniak’s (1995) study of the French speech community in Ontario shows that patterns of sociolinguistic stratification as described for monolingual societies such as NYC or Philadelphia can break down in bilingual speech communities which include marginal members whose language use, although grammatically adequate, is functionally restricted (in this case the marginal members of the speech community use French predominantly - if not exclusively - within the formal setting of the French medium school and only infrequently in informal conversations). Variation between the standard preposition *de* and the nonstandard preposition *à* (*l’auto de mon père vs. le char à mon père*), for example, which is socially stratified for core and semi-core speakers (i.e. speakers who use French frequently but not exclusively in informal situations), is absent from the speech of marginal speakers, who only make use of the standard variant *de*. Mougeon & Beniak have termed this process 'sociolectal reduction' (*ibid.* 92), i.e. the non-acquisition of socially significant variation by marginal members.2 Like Dorian’s semi-speakers and passive bilinguals, these 'restricted' speakers of French are nevertheless seen as members of the French speech community in Ontario.

Contemporary investigations of bilingual speech communities are relevant to the analysis of linguistic variation at the Cape. As discussed in § 2.2.1, Dutch immigrants never formed an

---

2 The loss of variation in such bilingual communities appears to be constrained by sociolinguistic salience, i.e. variables which show sharp stratification are likely to loose their non-standard variant, while patterns of gradual stratification often persist (Mougeon & Beniak 1995: 78). Apart from 'sociolectal reduction' the presence of marginal speakers in the community can also lead to the emergence of new cases of variation (typically grammatically 'simpler' variants: *ibid.*: 83-85, see also Thomason & Kaufman 1988: 51).
absolute majority at the Cape, and other linguistic groups such as Germans, French, Khoi and the ethnically heterogeneous slave population were gradually incorporated into the Netherlandic speech community, often maintaining (at least for a transition phase of language shift) their ancestral language. In the period under investigation (1880-1922) many members of the Netherlandic speech community at the Cape were bilingual in Dutch/Afrikaans and English, the politically dominant language. Unfortunately, the category 'language' was only included in the census from 1918, when the following percentages were reported for those classified as White: 42% bilingual, 29% English (only), 12% Dutch/Afrikaans (only). For children under 7 years the percentages were: 25% bilingual, 30% English, 45% Dutch/Afrikaans. This suggests that most of the bilingual adults come from originally Dutch/Afrikaans speaking homes, and thus entails the possibility of a language dominance shift (census data cited in Reinecke 1937: 597).

It is difficult to reconstruct to what extent English had invaded the informal domains in traditionally Dutch-speaking families in the 19th century, but a cursory look at, for example, the private and official correspondence of NGK ministers shows dominance of English in most private documents (diaries and private letters), while the official correspondence about church matters was regularly conducted in Dutch (see, for example, the papers by B.F.G. De la Bat NGK/ P-54/1 or T.C.B. Vlok NGK/P-24). A case in point is Andries Louw, son of ds Andries Adriaan Louw, who corresponded with his siblings and mother as well as with his fiancée and later wife Francina Malan exclusively in English. To his father, however, he wrote (with very few exceptions) in Dutch only. The same pattern of language use (i.e. Dutch to the father, English to everyone else) also characterises the private papers of his sister Maria Johanna. There is reason to suspect, that for many, English had become the dominant language of everyday life and especially informal peer-group conversations, while the use of Dutch was often limited to the religious and educational sector (for a more detailed discussion of the patterns of code choice and code switching see § 8.2). Bilingualism (Khoi-Dutch) was also characteristic of some communities, such as the Griquas (see § 2.2.1), who had settled in the Northern Cape. The Khoi languages found alongside varieties of Dutch in these communities were Gri (until the mid 19th century), Korana (until about 1900) and Nama (still spoken in many communities; see Traill 1995: 7-8). Amongst ex-slaves and their descendants Malay (see § 2.2.1) was spoken until the late nineteenth century (especially within the Muslim community; Franken 1953: 116ff., Valkhoff 1966: 261ff.). The Cape speech community thus parallels (within limits) the bilingual communities described by Horvath, Dorian and Mougeon & Beniak, i.e. speakers had degrees of membership in the Netherlandic speech community, which might have supported the non-acquisition of certain variants by marginal and semi-core speakers as well as the development of group-specific variants among more peripheral members, as observed by Horvath in Sydney.

A second major criticism levelled against the Labovian model of the speech community
focuses on the consensus view of society which underlies the model's emphasis on shared evaluation patterns (Williams 1992). Consensus models of society are rooted in functionalist social theory and stress the perception of society as a harmonious organism governed by an overarching consensus of values across social groups. Conflict-models of society, on the other hand, focus on the schisms in society and the divergence of interests and values between different social groups (for a discussion of the two approaches see Collins 1975, Waters 1991). Support for a conflict view of society and language is provided by Labov's more recent work on black and white vernaculars in Philadelphia, where he found sharp linguistic segregation between the two groups (Labov & Harris 1986, see also Rickford's 1986 discussion of his own work on sociolinguistic variation in Cane Walk). Furthermore, Milroy & Milroy (1992: 3) suggested that the persistence of non-standard varieties is probably best interpreted from a conflict perspective which assumes that different social groups have different norms and values. Support for the complementary use of both consensus and conflict models in sociolinguistics (which is generally seen as appropriate in sociology; see Dahrendorf's 1959: 163 'thesis of the two faces of society'; also Giddens 1997: 570-571) can be found in the description of gradient versus sharp stratification (Labov 1991 [1972]: 113-114). Gradient stratification (such as reported for postvocalic /r/ in NYC) can be interpreted as reflecting the absence of sharply conflicting norms across groups, while sharp stratification (as found for the dental fricatives in NYC) correlates with fundamental and long-standing class divisions (Guy 1988: 44).

The results of a recent study of linguistic variation in non-metropolitan Mexico (Santa Ana & Parodi 1998) indicate that not only the patterns of variation might reflect the existence of fundamental social cleavages, but also that the evaluation of variation is not necessarily homogeneous within a speech community. When conducting their interviews Santa Ana and Parodi noticed that 'some individuals, in contrast to the majority of interviewees, apparently demonstrated no awareness of any of the social evaluation patterns of language variation' (1998: 26). To account for these differences of evaluation Santa Ana and Parodi proposed 'a model of

---

3 Note that Labov was able to maintain a consensual view of the speech community in his Philadelphia study by postulating the existence of two different speech communities (one consisting of black speakers, one consisting of white speakers). The two communities were found to show different patterns of evaluation and different patterns of phonological and grammatical variation (Labov 1980).

4 Although the Milroys acknowledge the existence of conflicting norms on the level of the speech community, their Belfast study nevertheless assumes consensual norms within the class-based social networks (for a critical assessment of the Belfast research, see Williams 1992: 198).

5 The possibility that this could be explained by assuming the existence of two separate speech communities (as argued by Labov for Philadelphia, see fn. 3) was investigated by Santa Ana & Parodi, but was rejected since no separate linguistic feature could be found to characterise the out-group (1998: 26-27). The possibility of diglossia was also rejected as the speakers showed no recognition of a clear H/L
nested speech community configurations' (1998: 34) which allowed not only for the incorporation of the above cited findings, but also made it possible to combine the notion of the speech community with the interactional concept of social network (referring to the informal and formal relationships individuals maintain with other people, see Milroy 1980).

Those individuals who did not share the evaluative norms of the majority of the sample were marginal to the large-scale speech community, but nevertheless part of it. Santa Ana and Parodi summarised the linguistic behaviour and social position of these marginal members as follows:

[They] move in relatively restricted and close-knit social networks [...] limited to a handful of extended families and to very local interactions.
These speakers exhibit little conscious or unconscious recognition of differential language use of language by other speakers, and they minimally modify their speech to accommodate their interlocutor. [...] They may acknowledge some variation in speech; however, they are indifferent to the social judgements that are linked to such variation by other speakers. [...] Contacts with the socializing and evaluative social institutions that strongly affect people's sense of self and their speech, such as schooling and work outside the home, have been tenuous and brief. Other contacts with the larger social world are superficial, such as the passive reception of radio and television programs. [...] these individuals speak the regional dialect. They use non-standard lexical items and phonological forms, as well as stigmatized items of non-taboo content; but they do not demonstrate mindfulness of the reactions of other speakers to their language use. (1998: 38-39)

Extending the ties of the social network to include wider regional (as opposed to local) ties leads to a second group in the Mexican speech community, whose members recognise stigmatised variants and 'are sensitive about the way they speak with outsiders' (1998: 30-40). The network ties in this group are still dense and multiplex. Weak, uniplex ties characterise the third extension of the speech community which includes a large set of mutually anonymous speakers exhibiting significant class and educational differences. Members of this group recognise linguistic stereotypes and markers and are able to evaluate speakers on the basis of their language use. Finally, the largest group, encompassing the three groups discussed so far, is the national speech community (Santa Ana and Parodi termed the sub-groups speech locale, speech vicinity and speech district). Individuals who participate in this group are aware of the social prestige attached to the standard variety and usually aim at producing standard forms in their own speech.

The four groups indicated in this typology can be ordered on an implicational scale. As the individual speakers come to recognise and evaluate types of variables (stigmatised variants, regional variants, standard variants) they will obtain 'membership in a widening sequence of speech configurations' (1998: 41)
<table>
<thead>
<tr>
<th>stigmatised variants ('stereotypes')</th>
<th>regional variants ('markers')</th>
<th>standard forms</th>
</tr>
</thead>
<tbody>
<tr>
<td>I (locale)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>II (vicinity)</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>III (district)</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>IV (national com.)</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Table 4.1 A speech community typology (based on the degree of recognition of sociolinguistic norms; Santa Ana & Parodi 1998: 35)

This general typology developed for the description of non-urban settings is certainly more appropriate for the description of many historical speech communities than the conventional Labovian model, as it acknowledges (a) the fact that the supraregional standard variety and its norms might not be actively (or passively) known to a large number of individuals who are nevertheless members of the speech community, and (b) emphasises the importance of local norms and knowledge. The society at the Cape in the 19th century was certainly closer to rural Mexico than to NYC, and the Netherlandic speech community at the Cape can be defined loosely as including all those individuals who make use of a Netherlandic variety in at least some of their interactions. In terms of evaluation of the different variables it can be hypothesised that the situation might have resembled the one in Mexico, i.e. certain groups of individuals showed little awareness of the evaluative norms of the majority, and did not feel constrained in their use of stigmatised or regional forms, while others aspired (however ineffectively) to the Dutch standard. A considerable proportion of members of the Netherlandic speech community were probably bilingual, with some of them being more central members of the English (or Khoe/Malay) than the Dutch speech community.

The question of what constitutes an adequate sample size (i.e. the number of speakers to be sampled) has not yet been clearly answered for sociolinguistic research (for a critical discussion see Romaine 1980, 1982a: 109-110). Sociolinguists have sometimes used rather small samples: Milroy’s (1980) analysis of Belfast vernacular included forty-six speakers, Trudgill’s (1974) study of Norwich sixty speakers, Patrick’s (1995) study of Jamaican Creole fourteen speakers, and the investigation of variation in Mexican Spanish (discussed above) was based on interviews with thirty-five speakers. Sankoff (1980b: 51-52) has suggested that a sample of around 150 speakers is adequate even for complex and highly heterogeneous speech communities, while ‘samples of more than about 150 individuals tend to be redundant, bringing increasing data handling problems with diminishing analytical returns’. Small samples are, however, not necessarily unrepresentative and many important (qualitative and quantitative)

---

6 It should be noted that Sankoff’s remark is not based on statistical theory but on her experience with variation studies.
insights can be gained from the study of such samples. However, confirmatory data analysis (especially parametric tests of statistical significance such as ANOVA) should be used with caution on samples consisting of less than about fifty speakers and p-levels of > 0.01 should be interpreted with care (see STATISTICA 1995 I: 1413).  

The sample collected for this study consists of 136 individuals born and raised in various parts of the Cape Colony. Although the speakers grew up in the Cape Colony, some moved north in their adulthood, as employment prospects were better. This was, for example, the case for all of C.P. Hoogenhout's children. Nevertheless, the letters they wrote from Pretoria were included since all of them spent their linguistically formative years in the Cape Colony.  

4.1.2 Collecting the Data

Not all speech styles used by the members of a speech community are of equal interest for the study of structured variation and change. The 'vernacular', according to Labov a specific linguistic style used in informal conversations 'in which minimum attention is given to the monitoring of speech' (Labov 1991 [1972]: 208), is believed to offer the most systematic data for sociolinguistic analysis.  

To get access to the vernacular (and to minimise the observer's paradox, i.e. the distortion of the data as a result of the process of investigation itself) Labov developed the sociolinguistic interview in which the researcher tries to create an informal atmosphere by encouraging everyday conversation topics (for a comprehensive summary of the methodology see Labov 1984). To capture the individual's full linguistic repertoire Labov divided the interview into five 'styles', which are distinguished on the basis of setting (formal vs. informal) and which correlate to the degree of attention the speaker pays to his speech. He predicted a regular decrease of non-standard variants and high frequencies of (overt) prestige variants at the

---

7 It is not clear what qualifies as a small sample in statistical theory. According to the documentation of STATISTICA, the statistical package used in this study, small samples have N<50. However, other texts identify N<30 (Butler 1985: 55), or even N<10 (Ehrenberg 1986: 129) as a small sample.

8 There are only three exceptions: E.A.S. Schabort came from Bethlehem in the OFS, and Hester Hoogenhout (née Van der Bijl) and Hendrina Hoogenhout (née Viljoen) from the Transvaal Republic.

9 As Macaulay (1988) has pointed out the term 'vernacular' is not used consistently in the sociolinguistic literature. Apart from the above discussed variationist use of the term (i.e. referring to informal, casual speech), 'vernacular' can also refer to a non-standard variety used regularly by a particular geographical, social or ethnic group (such as, for example, AAVE in the United States), or an independent (usually relatively unstandardised) local language which is subjugated to a politically dominant lingua franca (as in 'the vernaculars of Africa' or the 'rise of the vernaculars' in Europe as national languages).

10 The approach was refined by Lesley Milroy in her work on Belfast (1980) where she used the anthropological method of participant observation to obtain samples of spontaneous speech.
formal end of the continuum. Casual speech (= vernacular) is indicated not only by (comparatively) low frequencies of prestige variants, but also by certain channel cues such as changes in pitch, volume or tempo (see Labov 1991 [1972]: 79ff.)\(^\text{11}\)

<table>
<thead>
<tr>
<th>style</th>
<th>attention</th>
<th>setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>casual speech ('the vernacular')</td>
<td>least</td>
<td>least formal</td>
</tr>
<tr>
<td>careful speech</td>
<td></td>
<td></td>
</tr>
<tr>
<td>reading passage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>reading word lists</td>
<td></td>
<td></td>
</tr>
<tr>
<td>reading minimal pairs</td>
<td>most</td>
<td>most formal</td>
</tr>
</tbody>
</table>

A major challenge for studies in historical sociolinguistics is to get as close as possible to the casual, spoken language of the time. That variation in the written language is worthy of sociolinguistic inquiry in its own right (Romaine 1982a: 16, Vachek 1989) is certainly correct; however, the question of whether and to what extent the sociolinguistic patterns observed in the written record parallel variation in the spoken language remains a vital question for historical studies (Lippi-Green 1994: 1-3, see also Kytö & Rissanen 1983: 473-475). There can be little doubt that many people write quite differently from the way they speak, and that the process of writing has a 'formalizing effect' on linguistic usage (Kytö & Rissanen 1983: 474)\(^\text{12}\). However, research since the early 1980s has shown that the differences between spoken and written language might have been overstated in the past. Recent comparative studies of spoken and written language registers have by and large failed to establish a clear boundary between the two channels, and patterns of linguistic variation were shown to be largely independent of medium (Besnier 1988, Biber 1988). Based on the results of such studies it has been argued that the assumption of a 'great divide' between writing and speech (which lies at the bottom of the bias of modern linguistics towards the spoken language) 'is increasingly less convincing once contextual and actual social uses of the channels are taken more fully into account' (Roberts & Street 1997: 168).

Today literacy is generally seen not as an autonomous, uniform technology, but as a skill whose organisation is shaped by the political, social and historical context in which it occurs, and to account adequately for the relationship between spoken and written language thus requires an ethnographic and sociohistorical perspective on literacy (see Basso's 1989 [1974]: 432 outline

\(^{11}\) The inclusion of reading on the same continuum as speaking has, however, been criticised, especially by British sociolinguists (Macaulay 1977, Milroy 1980, Traugott & Romaine 1985).

\(^{12}\) Maas (1988: 16) has argued that the 'strict linearity of speech forces all structuring activity on the short-term memory [...] [the] short-term memory is relieved during writing.' In other words, when writing speakers have access to linguistic forms which are stored in the long-term memory, and writing might thus facilitate the retrieval of less-used material or structures.
In the 19th century a public school system was established by the British colonial authorities, and with the inception of the position of the Superintendent General of Education in 1839, a centralised authority in educational matters was created (for more details see Malherbe 1925: 93 97, Pells 1938: 31-32). Although legislation was not yet in place, the school system was practically racially segregated from the second half of the 19th century. Public non-denominational schools which taught a wide range of subjects were frequented almost exclusively by White children, while the majority of Black and Coloured children received their education from mission schools, where instruction focused almost exclusively on the three r's of reading, writing and arithmetic (Bickford-Smith 1995: 25). Notwithstanding better access to educational institutions, schooling was still limited for many White children. According to the census of 1891, only 43% of White children of school-going age were enrolled at a school, and according to the report of the education commission daily attendance of those enrolled was irregular (Education Commission, Third and Final Report 1892: 4-6). Ten years earlier the so-called 'Ross Report' (compiled by Donald Ross who was sent out from Scotland to report on the educational system of the Cape Colony) summarised the low quality of instruction in the schools as follows:

In a large, by far the largest - number of schools it would be an absolute waste of time to examine the pupils in detail, so very elementary were their attainments and so inferior was the quality of instruction imparted. (1883: 4)

The quality of the educational system improved only after Dr Thomas Muir was appointed Superintendent General of Education in 1892. School inspections now took place regularly, educational surveys were conducted, curricula were extended, suitable textbooks were issued by the Department of Education and the training of teachers became a focal point of the department's activities (Malherbe 1925: 172, Pells 1938: 50, 94-95).

Not only was school attendance sporadic and the quality of teaching questionable, opportunities to practice the literacy skills acquired at school were not many in the colony. This was emphasised by Graff (1979: 310) for 19th century Canada:

There is little reason, in fact, to suspect that the daily culture of a nineteenth century city overemphasized the printed word, or that much literacy was required to learn its ways. Residence, commerce, and industry intermingled more than not; walking was sufficient to find one's way about. The city was a place of sights and sounds more than of print and text, with structures, both obvious and hidden, to be 'read' and explored with all of the senses.

Despite declining costs of production, books remained something of a luxury in the late 19th century with prices varying between two and ten shillings. Cheap books were generally badly bound and printed, and for the same amount of money one could have purchased five
pounds of bread, two pounds of meat or half a gallon of wine. 14 There is no evidence that anything comparable to the popular 'street literature' which existed in England and North America was available in the colony (on the genre of street literature and its importance for the development of literacy see Neuburg 1973: 206-107). At the Cape cheap and easily digestible reading matter was found primarily in newspapers and periodicals, however, compared to the sensationalist broadsides (the 'penny dreadfuls' and 'shilling shockers') the former constituted rather difficult and dry reading matter for the many semi-literates in the colony. Living conditions were often anything but conducive to the development of reading habits, and long hours of physical work left little time or inclination to exercise one's literacy skills. Although literacy remained peripheral for many, reading and writing were nevertheless deeply embedded in Cape society. Laws were no longer orally transmitted but written down, political changes and decisions were disseminated and publicised in writing, and the ability to manipulate pen and paper became increasingly important for individuals to protect and increase their social and political rights. Oral culture and oral practices interacted as elsewhere in the world with the emerging culture of the written word. In sum, considering the sociohistorical context of literacy, the 19th and early 20th centuries are of particular interest for historical sociolinguistics, since a growing school system enabled more individuals (including women and non-élite members of society) than ever before to acquire at least the basic skills of reading and writing, while at the same time schooling was too rudimentary for most to guarantee long-term exposure to standard norms.

Adopting Kytö & Rissanen's 'hypothesis of the closeness of informal speech to informal writing' (1983: 474, see figure 4.3), it was decided to focus in this study on personal documents only. The vast majority of texts are private letters to family members and friends; however, two private memoirs (M. Heroldt and A. Schabort) and excerpts from one diary (F.S. Malan) were also included. Some of the letters are only personal in the wider sense of the word, i.e. the addressee occupied a position of institutionalised authority in relation to the writer (typically brethren - minister/ missionary), yet, from the style and tone of the letter it is clear that the writer knew the addressee personally and that a relationship of trust existed. Thus, the informality of the texts can be described as a function of (i) the existence of a close relationship between writer and recipient, and (ii) a communication situation characterised by privacy (see also Rissanen 1986: 101).

---

14 See for example Catalogues van Boeken enz. (J. Dusseau & Co. 1897) and Prijslijst van Boeken bij die Paar Drukkers Maatschappij, Beperkt (October 1908), for general price and wage index see Burnton (1903: 234-326, see also Bickford-Smith 1995: 91-102, 168, 179).
For some individuals only a single letter could be located; for others several letters (sometimes addressed to different people) were available. All original documents were handwritten (ink or pencil) and no typed (or printed) material was included; the quality of the paper varies from heavy good-quality paper to cheap, yellowish note-book paper. Many of the writers were clearly inexperienced in the written medium exhibiting what historians have called the 'illiterate scrawl' (see Franken 1953: 179). However, texts were on the whole legible and morphological units (including inflectional endings) identifiable.\textsuperscript{15} Spelling conventions were highly variable in the documents. Most common were inconsistencies between -ij- and -y-, and variation between e and i, z and s, and final sch and s. Letter writing, as reflected by the sheer volume of some of the collections, was an important activity in the four decades between 1880 and 1922. The condolence letters sent to 'Katie' Malan (née Van Huyssteen) after the death of her husband in 1905, for example, fill several tightly packed boxes. The corpus includes intimate love-letters, affectionate notes between husband and wife, concerned enquiries and advice from parents to children, friendly gossip between siblings and friends, and letters asking a friend, parent or benefactor for help and support. The tone of most letters is similar to the spoken language, making use of a rather limited vocabulary, short sentences, a great proportion of first person references, exclamations and a fragmented style with idea units often strung together without connectives (see Tannen 1982, Chafe 1982). No attempt was made to collect additional

\textsuperscript{15} Regarding the pluralisation of nouns, it was sometimes impossible to distinguish between -s and -n terminals. Verb endings, on the other hand, were quite clear (i.e. -\(\theta\), -\(\iota\), or -\(e(n)\)), yet it was frequently impossible to determine whether a plural or infinitive form ended in -\(en\) or only -\(e\). The use of capital letters and punctuation was arbitrary in many documents and was standardised in the process of transcription, i.e. capitals were used for proper nouns and words at the beginning of a sentence, punctuation symbols were added in square brackets to clarify the syntactic structure. Formulaic expressions at the beginning and end of letters (such as \textit{Liebe XYZ} or \textit{liefehebende XYZ}) were excluded from the analysis. There were a few instances in which names (in particular surnames) were not clearly readable and suggested readings were put in square brackets.
formal letters written by the sampled individuals; thus the stylistic dimension emphasised by Labov was neglected.

When collecting the data a small but homogenous group of individuals was found to stand out in terms of their linguistic behaviour. Socially this group is characterised by the fact that all its members are highly educated and young (see §4.3.1 and § 4.3.2), while linguistically they are best characterised as bidialectal (or at least bi-literate) speakers, i.e. they make use of two distinct codes in their written documents which can be described relatively unambiguously as either 'Dutch' or 'Afrikaans' (for a detailed discussion see § 8.2.2). In the quantitative analysis the respective texts were considered separately and the individuals in this group were thus counted twice, such as F.S. Malan 1 ('Dutch' texts), F.S. Malan 2 ('Afrikaans' texts), which increased the number of individual cases from 136 to 152.16

The linguistic corpus comprises circa 130, 000 words, and between 100 and 6000 words were transcribed per individual.17 The texts were collected in the State Archives, the NGK Archives, the South African Library, the archives of the Moravian Church in Genadendal and the manuscript collection of the University of Cape Town. The boundary dates of the study are 1880 and 1922 (roughly two biological generations), with most of the texts written between 1890 and 1910. As described in detail in chapter three these years were characterised by public discussions of language issues and increasing attention to CDV or 'Afrikaans', culminating in the official recognition of Afrikaans as the third official language (alongside Dutch and English) in 1925 (in the following chapters the corpus is referenced as C-1880-1922).

---

16 Following DeBose (1992: 161) an utterance was considered to be in Afrikaans or Dutch 'to the extent that it is grammatically compatible with an autonomous description' of one or the other. In the case of this group of speakers it was possible to identify decontextualised samples from their letters uncontroversially as either 'Dutch' (maintaining, for example, inflected verbs and nominal gender in agreement with the norms of Dutch) or 'Afrikaans' (i.e. virtually complete loss of verbal inflection, nominal gender, use of the 'Afrikaans' pronoun variants and the 'double' negation) . Letters which showed code alternation (typically one or two paragraphs in 'Dutch', followed by one or two paragraphs in 'Afrikaans') were split for the analysis, i.e. the different parts were analysed separately (for further discussion see §8.2.2). The type of 'Afrikaans' used by these writers was highly focused and similar to modern standard Afrikaans.

17 In order to assess the relative size of the sample it can be compared to the Helsinki corpus for Early Modern English which (covering a period of roughly 200 years: 1500-1700) consists of 550,000 words (Nevalainen & Raumolin-Brunberg 1989: 167) and Raidt's 18th century corpus of CDV texts written by women on the basis of which she made a number of generalisations about language use and sex. Her corpus included 57 informants with texts ranging between 30 and 1000 words (1994 [1984]: 186).
Figure 4.4 Temporal distribution of the sample texts 1880-1922 (number of letters/diaries per year).

The social background of each individual was reconstructed on the basis of the information provided in the documents themselves, as well as information collected from other sources. Often the writers remarked on their socioeconomic circumstances in their letters, explaining, for example, why a debt hadn't been paid or asking for financial support from a friend or family member. In a letter to C.P. Hoogenhout [Mary] Basson described the economic circumstances of her family as follows:

wy leven zoo zuinig als wezen kan, van klederen koopen weten wy byna niet meer in de laaste paar jaren, want wy hebben geen inkommen hier en gy weet zelf wat voor voedsel uitgaat, ik verkoop niet eens meer een hoender [,] het wordt opgeslacht want wy zyn dikwyls in gebreken met vleesch (3/9/1906)\(^8\)

Death notices, if available, were an important source of extra-linguistic information. These notices give the place of birth and ethnic group of the deceased, the place and date of death, age upon death (which allows the approximate calculation of the year of birth), names of

\(^8\) Such remarks must, of course, be interpreted carefully and cannot be taken as direct evidence, since it is always possible that the writer exaggerated his situation to gain pity or financial assistance. However, by including different sources (death notices, deeds office register, voter's list) it was possible to assess (within limits) the reliability of the reports found in the letters and diaries.
surviving and/or deceased spouse(s), names of children and last residential address. Also given in most death notices was information on the legacy of the deceased (or this information was filed separately in the estate accounts), which gave an indication of social standing in terms of wealth and property. However, death notices were not filled in regularly, and many people died without their death being officially recorded. Other sources used for the collection of extralinguistic information were the voter's list for 1903 (indicating residential address, 'race' and occupation, men only) and the South African Directories (1883/4, 1903 and 1911) which list members of the municipal councils, ministers and individuals with business activities for each town or district (for a summary of the sources available for genealogical research in South Africa, see Lombard 1984). An interesting (though unintentional) characteristic of the sample is that subgroups in the data form relatively close-knit networks. However, since it was not possible to reconstruct networks for all individuals, the network variable was not used directly in the analysis (but see §4.3.2).

In sum, like all historical data the linguistic corpus used in this study is a 'relic' - a collection of texts which have survived into the present by accident, and thus constitutes what statisticians term a convenience sample. In other words, the composition of the sample is not so much the result of careful selection, but is largely determined by availability. Yet, despite these limitations, I believe that the corpus allows for certain careful generalisations about the structure of variation at the Cape, and following Nevalainen & Raumolin-Brunberg (1989: 100) one might say:

The selection will have amply fulfilled its purpose if it generates new textual and socio-historical hypotheses to be tested against larger corpora in the future.

4.2 The Linguistic Variables

The concept of the 'linguistic variable' is central to variationist, quantitative analysis. Briefly, a linguistic variable is a linguistic unit consisting of two or more identifiable variants which constitute social and/or stylistic alternatives. Their referential meaning is, however,

---

19 The calculation of the year of birth was problematic in that only few notices gave the date of death as well as the exact age (number of years and number of months). To achieve uniformity for all individuals only the number of years was used to estimate the year of birth (i.e. year of death minus age in years).

20 According to an estimate by one of the archivists of the State Archives in Cape Town death notices were only filled in for about 60% of the White population (J.F. Van der Merwe- p.c. May 1998). Furthermore, women were filed not under their maiden name but under the name of their husband, which made it impossible to trace some of the young female writers such as [Mimmie] Laubscher or Catharina E. Beyers (who were still unmarried at the time of writing their letters). However, their family background could often be assessed by locating the death notices of their parents.
identical (for the original definition see Labov 1966, for an overview Wolfram 1993). Most sociolinguistic research has focused on phonological variables, where the criterion of 'referential sameness' does not pose serious problems, and which have the additional advantage of high frequency. Whether linguistic variability of other sorts (morphological, syntactic, lexical, pragmatic/ discourse) is equally amenable to variation studies utilising the concept of the linguistic variable has, however, been questioned. While morphological variation (such as, for example, the absence or presence of a grammatical inflection as in he goes vs. he go) is generally seen as fulfilling the criterion of 'referential sameness', assumptions of syntactic or pragmatic equivalence as well as lexical synonymy have been controversial (see Lavandera 1978, Harris 1984, Winford 1984). Following Dines' (1980: 15) pragmatic approach, the variants of syntactic and lexical variables will be seen as being only 'weakly equivalent', but having nevertheless common semantic grounds.21

Since a central interest of this study is the reconstruction of the gradual stabilisation of 'Afrikaans' as a Netherlandic (standard) language morphologically and syntactically distinct from Dutch, the data will be analysed with regard to two main points of reference: 19th/early 20th century Dutch and 19th/20th century (standard and non-standard) Afrikaans.22 To use modern Dutch (and its dialects) as a point of reference is generally seen as acceptable since modern Dutch is relatively conservative, continuing many structures attested for Early Modern Dutch, which formed the superstrate in the early settlement (Lass 1990: 89, fn.). The following linguistic features were considered in the analysis:

I. morphosyntactic
   - loss of person and number distinctions in the present tense paradigm
   - loss of ablaut and -en in the past participle of strong verbs
   - loss of -t in the past participle of weak verbs
   - loss of the perfect/preterite distinction
   - loss of neuter gender (of nouns)
   - the emergence of the Afrikaans adjective inflection

21 Underlying these discussions is the question of the ontological status of linguistic variables, i.e. whether it is possible to utilise the linguistic variable as a mere analytical tool and procedural device (as suggested by Winford 1984, Wolfram 1991, Hudson 1996: 173), or whether the concept itself has to be seen as making (maybe untenable) assumptions about the very nature of linguistic variability (and thus language itself; Romaine 1980, Cheshire 1987).

22 If possible 19th century and early 20th century descriptions of Dutch were used in the discussion of the data (such as Talen, Kollewijn & Hettema 1908, Van Dale 1904, and for the norms of South African Dutch, Elffers 1893). However, reference was also made to more recent descriptions (De Voogys 1953, Donaldson 1981, Geerts et al. 1984, Shetter 1994). For Afrikaans, Donaldson (1993) and Ponelis (1993) were used as the main reference works, as well as Malherbe (1917).
II. morpholexical
personal, possessive, relative and demonstrative pronouns
the forms for the third person singular of 'to have'
the use of baie
the complementizer lat/laat

III. syntactic
generalisation of the Afrikaans infinitive with om.te
'double' or brace negation (nie ... nie)
use of objective vir
periphrastic possessive with se

Following post-Labovian work (especially Milroy 1980, Horvath 1985) scores were calculated not for pre-defined social groups but for individuals (see §1.2.2 for the epistemological assumptions underlying the two approaches).

Turning to the more technical issues of quantification, there is the central question of how many tokens (i.e. variants) of a variable are needed to be confident that the variation shown by a speaker is not random, but constitutes a general and reliable pattern indicating the speaker's variable competence. Unfortunately this question has not received much systematic discussion, and often quantitative results have been based on rather few occurrences. This is particularly true for non-phonological variables which generally occur with much lower frequency in a given stretch of linguistic data, and Mesthrie (1992), for example, worked successfully with frequencies between 1 and 14 tokens for relative clause variation.23 Romaine (1980: 190-93), who discusses the issue in some detail, emphasises that although research (Guy 1980: 20) has indicated that for phonological variables figures based on a minimum of 10 tokens per variable appear to be reliable (the ideal seems to lie around 30 tokens), it is not always possible to generalise this as a guideline.

In some cases 1000 occurrences may not be enough, while in others fewer than 10, and even as few as two, might show contrastive patterns of usage. (Romaine 1980: 192)

In this context the notion of 'salience' referring to socially stigmatised variants (stereotypes according to Labov's terminology, Labov 1991[1972]: 314-317) is important, and Dines (1980: 16) has argued that under such circumstances even a single occurrence of a variant can be 'socially diagnostic'. It is likely that the sociolinguistic salience of variables increases during processes of language standardisation, when certain variants are assigned to the 'new

---

23 To use frequencies as low as 1 defies, however, the objective of variation analysis, i.e. in order to detect variation we need a minimum of two occurrences. To use a small number of tokens per individual is not only problematic with respect to the question whether it is an adequate reflexion of the individual's linguistic competence, but also involves problems of data organisation (see fn.25 on the use of percentages).
norm' of the emerging standard language, while the non-selected variants are branded as 'vulgar' and 'dialectal' (see Stein 1994: 7-9). Considering the absence (and probably impossibility) of a general guideline for the number of tokens needed for reliable quantification, in combination with the fact that variants which belonged to the new system of 'Afrikaans' were openly discussed and often stigmatised from the 1870s onwards (see chapter 3), an arbitrary threshold of a minimum of six occurrences of a given variable per individual was defined (most individuals had however much higher token numbers, see appendix B.3). A result of this procedure is that the number of individuals available for the analysis of different variables is not stable, and only 36 individuals were included in the final multivariate analysis (i.e. they had more than six tokens for each variable; however a control analysis accepting frequencies as low as 2 included 81 individuals and replicated the results reported in chapter 8).

A second technical problem concerns the inclusion of variables (especially syntactic and lexical) for which the pool of possible variants has not yet been closed. Quantitative descriptions of such variables usually count their use in relation to some rather crude measure such as the total number of words, number of sentences, or hours of speech (Labov 1982: 87, Hudson 1996: 170). For the majority of variables used in this study the pool of variants was clearly limited, and only the use of objective vir was quantified per 1000 words (see § 7.2.3). The other variables were standardised as percentages, which made it possible to compare individuals regarding their use of Afrikaans variants (percentages were rounded to two digits). It should be noted that because of the often small numbers involved, the percentage values cannot be interpreted as being accurate to within one part in a hundred. In other words, the fraction 3 over 9 is rendered as 33%, but should be read as representing an approximation accurate only within roughly 5 percentage points.25

Anyone who has ever conducted quantitative work involving a large sample with literally tens of thousands of tokens knows that numerical mistakes are inevitable no matter how often the results are checked and re-checked. Despite the remaining possibility of error it is believed that the quantitative results presented in chapters 6-8 show reliable trends, and that possible counting errors do not affect the results in any major way.

---

24 This threshold roughly agrees with Scholfield (1991: 381) who suggested excluding cases that exhibit fewer than five occurrences.

25 Marsh (1988: 126) has drawn attention to the problems involved in the calculation of percentages from small samples, and has suggested that samples of less than 20 should not be rendered in percentages. However, since sociolinguists often cannot avoid working with small samples (especially in the context of historical linguistics, or when investigating syntactic variables), the practice of calculating percentages on the basis of samples < 20 seems to be generally accepted (see, for example, the data in Mesthrie 1992: 224-227, Patrick 1995: 342).
4.3 The Social (Extra-linguistic) Variables - Mapping the Social Universe

4.3.1 Age, Sex and Ethnicity

The success of sociolinguistic research depends not only on the proper specification and measurement of the dependent variable (i.e. language), but just as much on the meaningfulness of the social concepts which function as the independent (or explanatory) variable. While there is a general awareness that the specification of complex social variables (such as social class) can be difficult, the general idea of dividing the social world into compartmentalised units that are believed to form part of some social 'reality' is rarely questioned. In the epistemological discussions within the social sciences, however, the realist assumptions underlying variable analysis have received much criticism. The explanatory power of the correlational model was criticised harshly by Blumer (1969: 127-139), who argued that the social categories used in variable analysis do not possess a reality of their own but are mere referential shorthands, constructed through social and scientific discourse. To acknowledge that social (like linguistic) variables are theoretical constructs implies that there is no 'optimal' or 'correct' measurement scale for a given variable, since the selection of such a scale is nothing but the result of our conceptualisations. That one and the same variable can be measured on different scales is evident from everyday discourse:

We can and do swap freely from a categorical to an ordinal to a metric language in describing a concept like wealth or income. Thus, we might speak virtually in the same breath of 'the rich and the poor', 'those further up the pay ladder', and 'those fighting to maintain pay differentials.' Consequently we can claim with equal justification that the 'natural' organization of the concept was reflected by using dichotomous classificatory or by the use of a series of ordinal categories, or assuming metric measurement was apposite. (Pawson 1989: 46-47)

While it is possible to conceptualise the same variable on different scales, this is not to say that the identification of social variables is entirely arbitrary. The task at hand is to develop theoretically-grounded concepts which, though not having an 'objective' reality in themselves, can be seen as 'constructs well-founded in reality (cum fundamento in re)' (Bourdieu 1987: 5), and the choice between competing conceptions of, for example, 'social class' should thus not be

---

26 Cameron and Coates (1988: 13) have claimed that 'the main goal of quantitative studies has been the collection and analysis of data on linguistic variation; sociolinguists are not primarily concerned with the reasons for such variation, and their methodology is not usually designed to probe such issues'. However, approaching linguistic variation by correlating linguistic features (dependent variable) with social variables (independent variable) does imply a hypothesis of the explanation at hand. In other words, although sociolinguists have rarely spelled out their explanatory strategies, they nevertheless take it as a field-defining assumption that explanations are found in the realm of the social.

27 For a summary of the sociological critique see Pawson (1989); for a critique of the empiricist assumptions underlying Labovian sociolinguistics see Romaine (1982a: 274ff.).
made on the basis of their explanatory power (which is understood within the context of variable analysis as their ability to yield high correlations), but on the basis of their theoretical coherence (see Scott 1994: 935).\textsuperscript{28} Despite epistemological difficulties, variable analysis has been shown to be useful for the investigation of many sociological and sociolinguistic questions. The persistence of the empirical (statistical) associations between 'bits of language' and social characteristics (however crudely measured) must be seen as a major argument for the continued use of variable analysis in sociolinguistic research.

*The Variable 'Age':* Questions of conceptualisation and measurement also impinge on variables whose primary nature appears to be biological. Thus, the *prima facie* uncontroversial measurement of 'age' in years (or for young children in months) proves more problematic at closer inspection.\textsuperscript{29} While the measurement of age as a continuous variable (measured in years) is useful for biological (including psycholinguistic) analysis, for sociological purposes the more anthropological approach of 'life stages' is often considered more appropriate (Carter 1971:24, Eckert 1997: 152). In other words, few sociologists would argue that the social consequences of an age difference from 15 to 21 and a difference from 40 to 46 should be seen as equivalent, and to distinguish different 'life stages' is seen as reflecting the social consequences of age better than a continuous scale. Traditionally four stages are distinguished: childhood (*infantia*), adolescence (*adolescentia*), adulthood (*virilitas*) and old age (*senectus*; Mattheier 1987b: 79, Eckert 1997: 155-166).\textsuperscript{30}

The stage of adolescence is particularly important for the analysis of language change from below (i.e. below the level of awareness), and adolescents have been found to 'lead the age

\textsuperscript{28} This perspective contrasts with Labov's utilitarian approach to the conceptualisation of, for example, social class: 'Binary divisions into upper and lower class are of little value in sociolinguistic studies and conceal more information than they reveal. A useful view of the social distribution of a variable requires at least four divisions of the socioeconomic hierarchy.' (1990: 220). In other words, a specific class model is selected not because of its ability to account adequately for the social divisions in society (which might well be binary), but because of its goodness-of-fit with the observed patterns of linguistic variation (for a rather polemic critique of such an utilitarian approach see Pahl 1993).

\textsuperscript{29} The central importance of age as a social variable is based on the observation that a change in progress is usually characterised by a pattern of age grading. This was first demonstrated in Gauchat's pioneering study of the Swiss village of Charmey (1905: 205), and has since been shown repeatedly (for a detailed discussion see Labov 1994: 45ff.).

\textsuperscript{30} Such categories are, however, highly dependent on historical and cultural circumstances. Thus, it has been argued that 'childhood' as a separate phase of development did not exist in medieval times, when children were expected to act and behave like 'little adults'. Regarding more recent developments, sociologists have argued that children are now growing up so fast that the character of 'childhood' as a distinct phase is diminishing (see Giddens 1997: 38, Settersten & Mayer 1997).
spectrum in sound change and in the general use of vernacular variables' (Eckert 1997: 163). Within the stage of adulthood early and late adulthood are sometimes distinguished on the basis of differential social expectations, i.e. early adulthood is characterised by increasing financial and emotional independence from the parental home and the development of independent social relationships, whereas late adulthood is typified by the parental role and increasing social responsibilities. The stage of adulthood is of interest for sociolinguistic research as it is during this time that the individual (at least in contemporary, urban societies) typically comes into regular contact with the public and formal standard varieties of a language, and is expected to fulfill a large number of social roles (husband/wife, parent, employee, son/daughter, friend, adviser, etc.) which results in the acquisition of new forms of behaviour, including increasing awareness of the standard language and the standard language market (Mattheier 1987b: 80, Eckert 1997: 159).

Sociolinguists have measured age both on ordinal scales (e.g. distinguishing different age groups) and on continuous scales. However, to my knowledge there has been little discussion of the appropriateness of the choice of such scales from a sociological or biological-developmental perspective. In his study of Martha's Vineyard as well as in the survey of New York department stores, Labov (1991 [1972]: 22, 57f.) defined the youngest age group rather oddly as 14/15-30 years, thus grouping together adolescents and (young) adults. Similarly Milroy (1980) divided her data without much further discussion into two age groups (18-25 years; 40-55 years) which can be glossed intuitively as 'young' and 'middle-aged'. However, this grouping implies that age group 18-25 is equivalent in homogeneity to the age group 40-55, a hypothesis for which no further evidence is cited. In his more recent work Labov (1994: 60) has advocated an 'empirical approach to the question of age grouping'. He suggests that the 'continuous dimension of age' can be divided into a set of 'qualitative categories: ages 8-15, 16-19, 20-29, 30-39, 40-49, 50-59, 60-69, and 70 and over'. In a second step the strength of the relationship between these age groups and the linguistic variables(s) will be calculated by using multiple regression, and on the basis of the result larger age groups will be established (i.e. groups with similar correlation coefficients). Note, however, that the identification of the initial 'qualitative categories' is again arbitrary since no explicit sociological or developmental motivation is given.

For the survey of the Netherlandic speech community developmental aspects of the variable 'age' were marginal, as with only one exception the individuals sampled were post-adolescent at the time at which they wrote their letters or diaries (the exception is J. Malan who was about ten years old). The variable 'age' was thus considered primarily in terms of generation (or 'age cohort'), and two generations were distinguished:
• individuals born before 1865 (age group 1, about 42% of the sample), and
• individuals born in or after 1865 (age group 2, about 58% of the sample).

The rationale for this division lies in the sociohistorical circumstances: individuals born after 1865 clearly profited from an improving public school system (leading to greater exposure to the norms of both standard Dutch and English), and were exposed to a fully-fledged language nationalistic agenda in popular discourse; both of which were important sociohistorical developments with the potential to influence language use. Individuals born before 1865 were exposed to these ideas and discussions only during early adulthood, since Afrikaner nationalism only gained strength from the mid-1880s (see § 3.2).

The Variable 'Sex'/ 'Gender': The variable 'sex' has often been treated as mere biological category in variation studies (see, for example, the overview given in Hudson 1996: 193ff.). However, the recent rise of the term 'gender' indicating both a biological and a cultural category, shows that the issue is more complex than often anticipated. Recent gender research has shown that categories such as 'feminine' and 'masculine' are not static and unchangeable, but that each category can comprise different 'masculinities' or 'femininities', including gender identification across the biological divide. That gender is more than a mere biological category was also emphasised by Gal (1991), who has argued that gender identities are centrally shaped by the unequal distribution of power in society, and it was argued by Eckert & McConnell-Ginet (1992, also Wodak & Benke 1997) that sociolinguists need to be aware of the possibility of multiple and diverse 'gender' identities and gender 'roles' within a given society and should ideally develop an ethnographic perspective with regard to the question of gender identities. However, at the same time 'gender' is not exclusively socially constructed, and social scientists need to take care not to neglect the biological-evolutionary aspects of 'sex' that shape the sociohistorical instantiation of 'gender identities'.

The statistical information provided by the census reports suggests that at the Cape class and ethnicity interacted in important ways with gender. While, for example in 1911, 47 % of White women were classified as 'wives or daughters' exclusively occupied with household duties, only 12% of Black women (including Coloured and Indian women) were assigned to the same category. On the other hand, 43% of Black women were classified as holding an occupation in

It is not possible to maintain a sharp terminological distinction between 'sex' as a biological category, and 'gender' as a cultural category, as bodies, body images and biological processes are not simply biological, pre-cultural entities, but are (within limits) transformed and mediated through discourse and cultural norms (see Eckert & McConnell-Ginet 1992: 463-464, Wodak & Benke 1997: 129). However, the terminological distinction can be maintained in a weak sense, using the label 'sex' to emphasise the biological aspects and 'gender' to emphasise the cultural aspects.
the category 'agricultural' (which must have included a fair amount of 'household duties' of a more physical sort), as opposed to only 2% of White women. Furthermore, in the 1880s and 1890s more and more women (particularly from lower middle class backgrounds, primarily White or Coloured) entered the work force, thus challenging the social and moral values of Victorian society as such 'working women' found employment not only in the domestic sphere (as servants) but also entered in increasing numbers into employment in the newly emerging manufacturing sector (Ward 1991: 19). The assessment of diverse gender identities through detailed ethnographic analysis is certainly not beyond the scope of historical research; however, since our understanding of the Cape's social history is still rudimentary, an ethnographic description of gender and language use was outside the scope of this study. 'Sex' was conceptualised as a strictly biological variable only (64 female and 72 male).\(^{32}\)

The Variable 'Ethnicity': The term 'ethnicity' has by and large replaced 'race' in current sociological and sociolinguistic research. The term itself is based on the Greek noun ἔθνος which was used to refer to non-Hellenic tribes characterised by territorial and cultural/linguistic unity (Erikson 1987). Thus unlike 'race' which is usually thought to be biologically inherited and expressed by physical differences, the term 'ethnicity' (like 'gender') is used to emphasise cultural identity. It is, however, true that groups which share cultural traditions are frequently also biological kin, and modern definitions of ethnicity usually rely on a combination of biological and cultural criteria, i.e. common descent as well as a common history and culture (which can include language and religion, see Giddens 1997: 210).

At the turn of the century the vast majority of inhabitants of the Cape Colony were classified as African (Black; 59%), and lived mostly in the rural eastern territories of the colony (although migration to the cities started in the 1880s). Whites comprised 24% of the population, and those classified as Coloured (basically ex-slaves, Khoe and descendants of interethnic unions) 17% (based on the census report for 1904: 302-303).\(^{33}\) For the study of the Netherlandic speech community, individuals classified under the category African or Black are of marginal interest as these were by and large speakers of Bantu languages, and even in case of migration

---

\(^{32}\) According to the census of 1904, the population of the Cape Colony can be estimated at circa 1,870,000 people (the number is rounded and excludes dependent children younger than 14). Of these roughly 51% were male, 49% female (1904: 302-303).

\(^{33}\) The percentages for the different 'race' groups categorised and enumerated in the census reports remained stable for the period of this study. The census from 1891 gives the following figures: African 55%, White 25%, Coloured 20%. For 1911 the figures are: African 59%, White 23% and Coloured 18%.
to the cities were more likely to learn English than Dutch.  

Recent historical scholarship in South Africa has focused on the question to what extent the people thus categorised by the administrators developed analogous 'ethnic identities', and in this context the question of a Coloured identity has been subject of much controversy (for an overview see Bickford-Smith 1993). The administrator's category of Coloured is characterised by its diversity, including not only descendants of the heterogeneous slave population, but also the indigenous Khoi and as well as everyone who could not be classified neatly as being either Black or White (see Census 1904: xxi). Ross (1982) has argued that the specific nature of Cape slavery (characterised by the import of new slaves rather than biological reproduction and the distribution of slaves over isolated farms, see §2.2.1) inhibited the emergence of a Coloured culture and identity comparable to the 'slave culture' of the American South, embracing the slave population in its diversity. Thus, different ethnicities continued into the 19th century, and Coloured nationalism, which came into existence in the early 20th century, has been interpreted as an urban movement orchestrated by an educational élite which adopted the identity prescribed to them by the colonists but paid relatively little attention to its slave past. A rather different interpretation has been formulated by Adhikari (1992) who maintains that the common experience of slavery and subjugation was central to the emergence of a Coloured identity at the turn of the century. Thus, Coloured intellectuals did not simply adopt the racial categorisation imposed on them by the officials of the colonial government, but created their own identity on the foundation of a shared past.  

Whether the origin of Coloured identity is to be found in the history of slavery, the racial categories used by the administrators or a dialectical relationship between the two, the historical evidence shows clear signs of an inclusive Coloured identity from the early 20th century, cumulating in the foundation of the African Political Organization (APO) in 1902.  

Related to the discussions of the emergence of an embracing Coloured identity is the

---

34 This is exemplified by the case of Paulus Mangoea from Herschel in the Eastern Cape, who being relatively fluent in English - had to learn Dutch (around 1918) when he joined the NGK Sendingskerk as a missionary worker (see his letters in NGK S 52/56/1).

35 There is evidence of communities consisting of ex-slaves and Khoi (such as the Katriver settlement) in which such an embracing Coloured identity might have flourished prior to the 20th century (Adhikari 1992: 103-106). The origins of the Cape Town 'Coon Carnival' celebrating the traditional slave holiday also supports the argument that Coloured identity is rooted (at least to some extent) in the experience of slavery (Bickford Smith 1995: 188).

36 An important question is, whether this politicised Coloured ethnicity included those classified as Africans. Not only were their linguistic and cultural roots different, they were also concentrated in specific occupations (unskilled labour) and often separated residentially, thus preventing the development of an inclusive oppositional identity in the Cape Colony.
question of a separate Malay/Muslim identity (see Adhikari 1989). Apart from their traditional
dress and Islamic religious practices, the development of a separate identity (which was largely
limited to Cape Town) was also supported by the existence of strong occupational
('washermen', 'tailors' and 'fishermen') and residential ties (Bickford-Smith 1995: 35). The
Muslim community came to prominence in the Cape Town cemetery riots in 1885/86 (which
erupted after the city authorities closed the Muslim cemeteries on Signal Hill), and in the late
1880s members of the Malay/Muslim élite put forward their own parliamentary candidate
(Bickford-Smith 1995: 195ff.). However, since Malays/Muslims 'shared similar experiences
of racial exclusion and marginality with other proto-coloured groups' (Adhikari 1989: 10) their
incorporation into a broader Coloured identity was easy.

In the early 20th century attempts were also made by some mission churches to
incorporate those classified Coloured as marginal members into the Afrikaner community, in
order to prevent the development of an independent and oppositional identity. Such a position
was advocated in De Ebenezer, a religious monthly paper edited by F.N. Van Niekerk (1911-
1923 available in the SAL). An inclusive language- and religion-based definition of Afrikaner
nationalism was also supported by the Moravian paper De Bode (see § 3.1).

White ethnicity was no less heterogeneous than other 'ethnicities'. Although public
discourse emphasised shared European ancestry and cultural heritage as a uniting factor, both
English and Afrikaner ethnicity (defined primarily in terms of language and religion) were on
the rise in the late 19th century (for Afrikaner ethnicity see § 3.2, for English ethnicity see
Bickford-Smith 1995). There are no census data for the number of English and Dutch/Afrikaans
speaking Whites, but Watts (1976: 43-44) estimated for 1890/91 248,000 speakers of English,
and 372,000 speakers of Dutch/Afrikaans.

The neat racial classifications of the census reports conceal the fact that by the late 19th
century the conceptual borderlines separating the 'races' were blurring, and intermarriages were
taking place with increasing frequency (especially among the working classes).

We have been astonished and from the standpoint of our social prospects disheartened, to find
how surprisingly large is the percentage of such settlers who marry coloured women. Let the

37 In the 19th century the term 'Malay' was used as a synonym for 'Muslim', and no longer referred
to a separate ethnic group. The majority of Muslims were classified Coloured and of diverse ethnic
backgrounds (primarily ex-slaves). Only 31 Muslims were classified White in the census of 1891 (xxxvi,
i.e. less than 1%).

38 In his 'working paper' on Malay/Muslim ethnicity, Adhikari (1989) argued, however, that these
episodes were not examples of exclusive Malay/Muslim resistance as both Malay/Muslim and other
Coloured groups were involved in the riots. Effendi (the Malay/Muslim candidate for parliament) also not
only emphasised Malay/Muslim interest in his campaign but 'was careful to project himself as a working
Imperial Government take a Census on this point in, say, District Six of Cape Town, and the result will astonish them. (South African News 4.4.1901, quoted in Bickford-Smith 1991: 27 see also Bickford-Smith 1995: 127f.)

A related matter is what has been described as 'passing for white', i.e. the upward mobility of Coloured individuals whose complexion was 'pale enough' not to prevent their participation in the upper ranks of colonial society and institutions usually reserved for people of European origin. Policies of increasing 'racial' segregation (occupational, educational and residential) were however initiated from the 1890s. Segregation was advocated not only as a solution to what became known as the 'poor White problem' (referring to the increasing numbers of unemployed Whites), but also served to fulfill the sanitary desires of upper class Whites 'to be protected from contamination by 'dirty' Blacks' (Bickford-Smith 1995: 102), and satisfied the capitalist's need for an available and controllable African labour force.

Although the variable ethnicity (in both its biological and its cultural dimensions) is of great importance for a thorough understanding of the sociolinguistics of the Netherlandic speech community, and an attempt was made to collect data for Coloured speakers in the archives of the Moravian Church and the Sendingskerk archives, the decision to include only personal documents prevented much material from inclusion in the corpus, which as a result is dominated by White individuals (120 out of 136). Further archival work should aim at closing this gap. This will, however, only be possible within limits as literacy itself was racially skewed (according to the 1891 census 90% of Whites older than 15 years claimed to be able to read and write, as opposed to about 15-20% of those classified Coloured).

4.3.2 Social Class

While age, sex/gender and ethnicity have been shown to be important aspects of social structure, it has been argued that socioeconomic class is the most pervasive category affecting all other aspects of social life and identity (Milliband 1987). Yet, of all social variables commonly used in variable analysis, the concept of 'social class' is probably most centrally in need of clarification since its theoretical foundations (most generally, production and market

---

39 Abdullah Abdurahman, for example, who became later president of the APO, was educated at the South African College School (SACS) which was officially for 'Whites only' (Bickford-Smith 1995: 143, see also ibid. 24).

40 While Whites have always formed a significant part of the poor, it was only in the late 19th century that poverty became defined in racial terms, i.e. the Victorian distinction between the deserving and undeserving poor was now drawn along racial lines (Bickford-Smith 1995: 126-127).
relations) have rarely been spelled out in sociolinguistic research. Most early sociolinguistic studies have conceptualised social class in terms of gradient status hierarchies, an approach usually associated with functionalist social theory and a consensus view of social structure (see Gouldner 1971, Crompton 1993/1996). Such continuous status hierarchies are established on the basis of occupational prestige (established on the basis of subjective rankings), income, housing and education (multiple indicator model), but tend to neglect structural categories such as property or production-market relations which have been central in conflict approaches to social structure. As discussed in § 4.1.1 with reference to the notion of the 'speech community', the consensus model underlying the use of status hierarchies has since the late 1980s been the object of serious criticisms, and conflict models have gained popularity. However, except for Rickford (1986), Guy (1988) and Milroy & Milroy (1992), class analysis within sociolinguistics has rarely drawn on the traditions of social theory, and the following overview is thus meant to provide the sociological background to the class scheme used in the analysis.

Class Analysis and Social Theory: The class analysis of Karl Marx (1818-83) can be seen as forming the nucleus of conflict models of society. Marx's interpretation of social structure and social change is based on what he has called the materialist conception of history, i.e. the economic conditions of production (base structures) are seen as the main factor determining the structure of all other areas of social, political and cultural life (superstructures) (as outlined in Grundrisse 1973 [1857-1858]). According to their relationship to the means of production, i.e. the resources necessary to produce goods (such as natural resources, factories and machinery), two principal classes are identified: the bourgeoisie or owner class and the proletariat or non-owner/producer class. As is well known, the relationship between the two classes is characterised by domination and progressive exploitation: in a competitive economy the owners of the means of production are constrained to 'exploit' the proletariat in order to make profits which are necessary for their economic survival and the maintenance of their dominant social position. The proletariat, on the other hand, being compelled to sell their labour under these exploitative conditions, will try to end the domination of the bourgeoisie. The economic and social interests of the two classes are thus diametrically opposed, leading to a deep-rooted and increasingly antagonistic class conflict which constitutes the major motor for social change. The conflict between the classes is not limited to the economic sphere, but extends to the realm of ideas and culture (including language), and based on their position within the relations of production each

---

41 It is thus notable that there is no chapter on 'class' in the recently (1997) published Handbook of Sociolinguistics, which features chapters on 'gender' and 'age' (see also the reviews by Cameron 1998, Horvath 1998).
social class will develop its distinct culture, values and ideas. Effective political and social action which will, according to Marx's political vision, eventually lead to the proletarian revolution, depends on the proletariat's developing class consciousness, i.e. the proletariat needs to become aware of its structural position within the capitalist system. A major problem of traditional Marxist class analysis is his commitment to a dual class model which does not allow for the conceptualisation of a stable middle class. Although Marx (1962 [1894]: 862) recognised the existence of intermediate classes, he saw these classes as standing outside the fundamental conflict between bourgeoisie and proletariat, and thus of no relevance to his theory of society and social change. The rise of the middle classes to an independent and pivotal part of society was not anticipated by Marx, and for neo-Marxist class analysis the preoccupation has therefore been the location of these intermediate or middle classes whose class position is ambiguous since they are both 'exploiters' and 'exploited' (see for example Poulantzas 1984 [1973], Wright 1985).

Max Weber's (1864-1920) thinking on social stratification, although strongly influenced by Marx, was critical of some of Marx's views, in particular his materialist conception of history and human relations. According to Weber, non-economic factors such as ideas, sentiments and intentions play an independent role in the development of society, and inequality is grounded in a multiplicity of structural determinants (property/marketable skills, prestige, political power) which lead to a variety of conflicts and differential processes of stratification. Weber's approach to the analysis of social structure is thus pluralistic, and based on three independent categories: class, status and party.

Weber agrees with Marx in that social class is determined by the economic structure, but argues that economic factors other than property account for the patterns of inequality and domination, such as sought-after skills and qualifications which determine one's ability to command income, wealth and other resources. The central concept in Weber's analysis is the market where skills and property are exchanged, and the class situation of individuals is ultimately determined by the marketability of their property and/or skills (Weber 1956 [1921]: 532). Based on the conception of classes as groups of individuals with similar economic interests

---

42 Note that for Marx, classes are objective causal realities and not mere theoretical constructs. This 'realist' position was defended more recently by Lee (1994).

43 For Marx 'class consciousness' is not the same as psychological consciousness (i.e. subjective identification with a group) as analysed in, for example, social identity theory (for social identity theory see Tajfel & Turner 1979, also Hogg 1992: 90ff). Marx introduced in his writings a new, anti-Hegelian meaning of consciousness, in which consciousness is determined solely by the materialistic conditions the individual lives in. Class consciousness is, thus, not what a single proletarian (or even the proletariat as a whole) regards as its aim at a certain point in time, but the ideas, feelings etc. which people would have if they were entirely aware and rational about their 'objective' class situation (see, for example, Marx 1971, 1: 705).
and opportunities in the market, Weber identified four different social classes: the working class (manual labourers with varying qualifications), the petty bourgeoisie (self-employed farmers and craftsmen), the propertyless intelligentsia and specialists (white collar workers, technicians), and classes privileged through property and education (rentiers, entrepreneurs, professionals with sought after expertise) (Weber 1947 [1922]: 427, 1956 [1921]: 179). Besides class, which is economically motivated, Weber distinguishes status (Stand) and party (Partei) as two other aspects of social structure. Status differentiation is based on the social honour or prestige assigned to certain groups by others, and is typically determined on the basis of criteria such as education, hereditary as well as occupational prestige and, most importantly, life styles and consumption patterns (Weber 1956: 538). Status differentiation is independent of class divisions and status groups can be formed across class lines. In other words, people in the same class situation might share very different status positions. For example, the social status of immigrant workers is usually lower than the status of national workers (regarding the interpretation of ethnicity and sex/gender in terms of status, see also Scott 1994: 940). An important difference between classes and status groups is that the latter always form communities (Gemeinschaften). The development of class consciousness, on the other hand, is contingent, depending upon particular sociohistorical conditions. Status groups typically evolve into a close caste and higher status groups operate mechanisms of social closure to keep out those below (Weber 1956 [1921]: 23-24). Weber's third category, party, refers to political factions and interest groups. Parties can cut across both class and status differences, such as for example, nationalist or religious movements, which often unite followers from different social backgrounds (1956 [1921]: 539).

Bourdieu's recent approach to the analysis of social structure has been described as an integration of a predominantly Marxist perspective on class (economically determined) with Weber's concept of status groups characterised by social prestige as well as specific consumption patterns (Brubaker 1985: 747). Bourdieu conceives of the social structure as a multidimensional space in which individuals are positioned according to the different forms of capital (and thus sources of power) they control. Capital, according to Bourdieu, exists in both material (economic, monetary) and immaterial forms, and under certain conditions these different forms of capital can be converted into one another. Four main forms of capital/power can be distinguished (note that the last two incorporate aspects of what Weber describes as 'status', thus for Bourdieu social 'status' is a form of capital):

---

44 This is made explicit in the preface to Distinction (Bourdieu 1984[1979]: xi-xii): 'The model of the relationships between the universe of economic and social conditions and the universe of life-styles which is put forward here [is] based on an endeavour to rethink Max Weber's opposition between class and Stand.'
I. Economic capital, which includes property, wealth and financial assets.

II. Cultural capital, which includes knowledge, skills and differing types of (recognised) educational qualifications (including linguistic skills and resources such as knowledge of the standard variety).^{45}

III. Social capital, which refers to resources 'based on connections and group membership'.

IV. Symbolic capital, i.e. the prestige and reputation associated with a person or position as a result of acknowledging the other forms of capital as legitimate (see Bourdieu 1985: 724/ 1987: 4).

Individuals are distributed in the social space according to (a) the total amount of capital they possess, (b) the composition of their capital (i.e. the relative weight of the different types of capital), and (c) their trajectory in the social space, i.e. how their initial capital is likely to be transformed throughout life histories (Bourdieu 1984 [1979]: 114, 128-29).^{46} The closer the relative position of individuals in the social space (i.e. the more similar they are with regard to capital ownership) the more likely their participation in shared social practices. In other words, occupants of similar positions in the social space are also subjected to similar conditionings which lead to the development of similar dispositions, tastes and behaviours (1984[1979]: 725). Bourdieu's monumental analysis of French society in Distinction ([1979] 1984) offers a detailed and empirically based account of the interconnectedness of class position (constructed in terms of occupation) and development of class-specific social practices (the so-called class 'habitus').

Social Classes and Class Conflict in the Cape: While a general understanding of social theory and class analysis is vital for the social historian of language, it is equally important to pay attention to the historical peculiarities since 'class' is not a universal, a-historical category, but is (just as gender and ethnicity) constructed through social practice and discourse, as well as through historically specific market and production conditions.

Nineteenth-century government statistics reflect a general interest in the socioeconomic classification of the population and a variety of economic classes and sub-classes are usually

---

^{45} While Bourdieu has largely focused on institutionalised forms of cultural capital (i.e. recognised educational qualifications, the standard language, artistic traditions accepted by the dominant culture), Thornton (1997: 200-209) has extended the concept to include what she terms 'subcultural capital', i.e. modes of knowledge and speech which operate outside the dominant culture (that is on alternative markets, see § 3.5), and which are not universally recognised as 'capital' (essentially what sociolinguists have described with reference to linguistic variability as 'covert prestige'). As Thornton defines it: 'Subcultural capital confers status on its owner in the eyes of the beholder.' (1997: 202).

^{46} Bourdieu's concept of 'class' has been criticised for its generality: 'class ceases to designate (as it does in Marx and Weber) a particular mode of social grouping: it becomes a metaphor for the total set of social determinants' (Brubaker 1985: 769). Thus, for Bourdieu, age, sex and ethnicity are not social dimensions interacting with class, but they themselves constitute class divisions, albeit secondary ones (Bourdieu 1984 [1979]: 103).
distinguished in the census. The classification of occupational groups in the Cape Census reports (1875, 1891, 1904 and 1911) is, however, of little use for a sociological analysis of the class structure, as people are classified not according to class position (understood in terms of capital owned and market position) but with respect to the economic sector they work in (i.e. domestic sector, commercial sector, agricultural sector etc.), and individuals grouped together under for example 'mercantile persons' (a suborder of 'commercial class') embrace a variety of market positions (in terms of skills) and property relations.

Merchants, Wholesale and Retail Dealers, all Speculators, Traders, Hawkers, Pedlars, Tuggers, all Capitalists, Financiers, Bankers, Bank Officers, Joint Stock, including Insurance, Company Managers and their Clerks, all Commercial Agents, Travellers, Commercial Assistants, Clerks, Salesmen, all Auctioneers and Valuers. (Census 1891: ixv)

While some of the merchants might have owned property and accumulated capital, the same is most certainly not true for hawkers and pedlars (Bickford-Smith 1995: 17). Within the agricultural sector, socioeconomic differences were also much more divisive than suggested by the all-embracing category 'people possessing, working or cultivating land/or animals' which accounted for over one third of the population.47 There was the wealthy Cape gentry which consolidated its power largely in the 18th century, 'almost inevitably white and in general considerable employers of labour' (Ross 1981: 1), there were small and middle-sized independent farmers of different ethnic origins, from well-off to poor, there were the so-called bywoners (farm tenants who were given the use of land in exchange for either a share of the crop or seasonal labour services) and farm labourers.

The manufacturing sector in the Cape was still 'embryonic' at the turn of the century, although the first factories emerged in the 1880s (Ward 1991). This meant that unlike in 19th century England or Germany, the 'industrial proletariat' was small in numbers (Bickford-Smith 1995: 37). Many people worked, however, as casual, unskilled labourers, either on farms or in the cities. The casual labour force at the Cape was very heterogenous, including not only urban low-skilled and low-paid wage earners, but also the landless rural poor, many of whom tried for better employment opportunities in the cities (Bundy 1984). Skilled workers, such as the railway workers in Salt River, and artisans, were in a slightly better position, and some managed to enter permanent employment, or were able to establish their own small workshops (Bickford-Smith 1995: 21). Although the pool of manual workers was heterogeneous in terms of ethnicity and

47 In 1891 44% were working in the agricultural sector (1904 42% and 1911 38%). As a result of the prominence of the agricultural sector, the majority of White and Coloured inhabitants lived in rural areas, however, urbanisation was already considerable (in 1891 41% of Whites and 40% of Coloureds lived in urban areas).
skills, spontaneous strike actions (including desertion and boycotting) and growing trade unionism from the 1890s onwards point to the emergence of a group identity as well as the existence of social conflict (Mabin 1983, Bickford-Smith 1981, 1991, 1995: 96-9, 108-9). The increasing polarisation of the interests of employers and workers was reflected in the Masters and Servants Act from 1873, which specified penalties (ranging from £1 to £5 and/or imprisonment up to three months) for employees. 'Careless work', 'neglect', 'refusal to obey' and 'bad language' were specified in the act as sufficient for severe punishment - in short, 'virtually any behaviour [...] that threatened an employer's authority' (Bickford-Smith 1995: 94).

The fact that most manual work was casual employment brought with it that the boundary between manual workers (including many artisans) and the unemployed poor was practically non-existent. People in the cities as well as in the rural areas drifted in and out of employment, and the plight of the pool of casual labourers and their families was poverty, bad sanitation, shortage of housing and high rates of mortality. For members of this group upward social mobility was unlikely, and several economic depressions in the second half of the 19th century, the Anglo-Boer War and the increasing mechanisation of agriculture all contributed to an increase of poverty after the turn of the century. At the opposite end of the social scale was the 'dominant class' or 'grand bourgeoisie', predominantly English-speaking and including the owners of great mercantile firms in the cities, many of whom owned shares in the gold and diamond mines in the north. Members of the landed gentry and leading civil servants (with incomes of over £1000 a year) as well as some professionals also belonged to this group, which was close-knit and exercised considerable political and economic power (Bickford-Smith 1995: 20). An emerging privileged class consisted of white collar professionals such as doctors, lawyers, teachers and ministers, i.e. members of the educational élite, most of whom possessed few liquid assets or property but scored high on 'cultural capital' which (given the right circumstances) was convertible into 'economic capital' and/or 'symbolic capital' (for a general discussion of the rise of the professional classes from the late 19th century see Perkin 1989, see also Hofmeyer 1987: 93). Note also that economic and cultural assets had an important impact on the distribution of political power in the Cape Province as political rights were limited to literate property owners.48

The class scheme used in this study is eclectic, combining influences from Marxist and Weberian class analysis with the recent reworking of the models by Bourdieu (especially his forms of capital, i.e. the notion of cultural capital as a form of power). The dismissal of the functionalist's consensus view of society is motivated by the historical reality in the Cape Colony

48 The franchise qualifications (municipal and provincial) were altered several times in the late 19th century, always in the direction of reducing the popular vote. In 1892 the Franchise and Ballot Act raised the property qualification from £25 to £75 (thus dis-enfranchising many artisans) and voters had to be able to sign their name (Bickford-Smith 1995: 144). According to the census of 1904 69% of White and 41% of Coloured and Black men qualified for the vote.
at the turn of the century, where rapid industrialisation and accelerated socioeconomic change led to social polarisation and conflict (as documented in detail in Bickford-Smith's work on the emergence of bourgeois attitudes towards the 'poor' and increasing residential and social segregation in Cape Town). Following the tradition of empirical sociology (see Crompton 1996: 117) occupation is seen as a proxy for class position, indicating the individual's position in the market. The advantage of employment based class schemes is that occupational information is not only readily available and easily interpretable, but can also be compared with official census figures and other studies. Class position was, however, not assigned solely in terms of occupation, but also on the basis of other information regarding financial or property assets (obtained from death notices and estate accounts). The class model includes the following classes:

I The Dominant Class (Grand Bourgeoisie)
Owners and managers of large amounts of economic capital and/or property; high ranking politicians; top civil servants; some professionals.
Class position is determined by the possession of high amounts of economic capital, as well as by significant amounts of social/ symbolic/ cultural capital.

II The 'New' Professional Class
The majority of civil servants, clerical workers, teachers, and other professionals. Often propertyless but also including owners of some property (usually urban).
Class position is determined primarily by the possession of cultural/ symbolic capital; geographical and social mobility is common.

III The Middle Classes (The Petty Bourgeoisie)
a. Independent farmers owning moderate amounts of land but not commanding much disposable economic capital. Employers of labour. Class position is determined by property holdings, but small amounts of other forms of economic capital. Low on cultural capital, but relatively high on symbolic capital (often occupying a position of authority in the area; deeply integrated into close-knit local networks).
b. Small scale production and ownership (including shop owners and some economically successful independent artisans). Small to moderate property owners, but disposable economic capital and average/moderate amounts of cultural capital. Often deeply integrated into close-knit local networks.

IV The Working Classes
a. Most artisans, skilled manual workers and domestic servants. Small-scale farmers and bywomens. Marketable skills, but generally low on all forms of capital.
No or extremely small amounts of any form of capital.

The classification should not be regarded as a definite 'map' of class structure; it is no more than a (rather crude and tentative) *instrument du travail* which can help to organise and structure the socioeconomic dimensions of the data set in a way that is informed by social theorising.
At this point a few comments on the relationship between class and social network are worthwhile. Milroy & Milroy (1992) have argued that weak ties are typical for those sectors of society where social mobility is common, while (at least in modern, urban societies) the working class as well as the dominant class (the power élite) usually maintain dense and close-knit networks. Following Milroy & Milroy (1985b, 1992), the dominance of weak ties among members of the geographically and socially mobile new professional class has two important consequences for language change:

(i) The loose-knit, multi-tie networks of this group 'ensure that the dominant linguistic market - as embodied in some form of legitimized or standard language - holds sway without hinderance from [...] alternative linguistic markets.' (Milroy & Milroy 1992: 22)

(ii) The loosening of the norm-enforcement mechanisms connected with the increase of weak ties makes this group particularly susceptible to innovations, and their weak-tie network structure facilitates their diffusion (see § 8.1.4).

Unfortunately, the census data do not allow for detailed statistics regarding the quantitative dimensions of class structure since the occupational categories used in the census reports give on the whole little indication of the market/property situation of the individuals thus classified. Nevertheless an attempt was made to reclassify the occupational information provided by the census in order to arrive at a general estimate of the respective size of the different classes in the Cape society. Following Stedman Jones (1976: 352) numbers given for very general and embracing headings were placed in the social class which pertained to the majority of occupations listed under that heading (children under the age of 14, scholars and 'wives and daughters' were excluded). Individuals grouped under the category 'farmers' were seen as representing by and large class IIIa, while the landed gentry were more likely to be indicated under categories such as 'landed proprietor', 'persons of property or rank' and even 'capitalist, financier'. A final problem was posed by the category 'peasants'. It appeared that this category referred primarily to the African farmers in the Eastern territories, which were incorporated into the Colony after 1875. Since this group does not really form part of the colonial society in which the Netherlandic speech community was situated, but rather constitutes a separate (politically marginal and subordinated) sociolinguistic entity, it was excluded from the quantitative estimate. Percentages were rounded and should be seen only as very general estimates. Based on the

---

48 For example, the occupational category 'builder, building contractor, foreman, measurer, clerk' (census 1905: cxxxix) embraces a number of rather different class and market positions, mixing manual and non-manual occupations, different skills and probably also differences in the amount of economic capital possessed (regarding the lack of clarity in the census categories see also Bickford-Smith 1995: 20).
information provided by the census data one can safely assume that class I comprised only a minority of the population (about 2%). Judging from the census reports class II was small but growing and can be estimated at about 10% of the population. Both, class I and II were almost exclusively White (close to a 100% for class I, and about 80% for class II). Class III was also predominantly white (about 70%), and can be estimated at about 25-30% of the population. Class IVa was ethnically more heterogeneous, as many of the artisans were ex-slaves (or descendents of slaves) and had acquired their skills during the times of bondage. For example, of the over 8000 masons in the colony, more than 5000 were non-White (1904: cxlv), as well as about 2000 of the over 9000 carpenters and joiners (1905: cxi). The position of this class is ambiguous in that its members typically sought distinction from the lower stratum working class and the poor, but at the same time rejected patronage from the upper classes, emphasising a separate identity. At the same time the nature of employment in the colony, which was largely casual, blurred the distinction between artisans/skilled labourers and unskilled labourers (IV b, see above). Using these figures as a general guideline the following general distribution of social classes in the Cape was estimated (very roughly) at: class I 2%, class II 10%, class III 25%, class IV 63%.

Figure 4.5 Approximation of the general distribution of the social classes at the Cape (in percentages; based on the census 1904, excluding the category 'peasants').

The class membership of the individuals in the sample was assigned on the basis of occupational information as well as financial and educational assets. In agreement with sociological thinking, children and adolescents were categorised according to the class position of the parental household. In the case of women class position was assigned on the basis of the father's or husband's position if the women had no independent employment; however, women
who had employment were classified individually. Although the classification was fairly straightforward for the majority of individuals, there were also a number of problematic cases. For example, D.G. Malan started out as a teacher in a small school in the Boland, but later bought a farm in Paarl and became a farmer. Should he be assigned to class II ('new' professional class) or class III? Or Louis Botha? Should he be classified as a member of class III (farmer) or rather as a member of class I (politician)? It was decided to use in such cases the first occupation someone had, i.e. D.G. Malan was classified as a member of class II, Botha as a member of class III. Looking at the overall distribution of the sample across the different social categories, it is not surprising that the result is not representative of the Cape society at large, and that certain groups (i.e. class II and III) are clearly overrepresented. This, however, was unavoidable since a social (and for that matter ethnic) bias existed in the distribution of literacy, which was by and large limited to classes II and III.

Figure 4.6 The distribution of the sample across the social classes (in percentages).

4.4 Summary

In order to assess the role social factors play in the patterning of linguistic variation carefully stratified corpora are necessary. Variable analysis, although 'crude' and often inadequate for an understanding of what people 'do' with language, is an important starting point which allows us to generate hypotheses, which can then be tested in small-scale analyses focusing on the context-bound agentivity of individuals. The basic assumption underlying variable analysis is that similarities in the social location of individuals give rise to group-specific practices whose regularity can be detected by means of (descriptive) statistical techniques. Social variables such
as class or ethnicity generally fall into the conceptual domain of 'resources', that is, 'structured properties of social systems' (Giddens 1986: 15, see § 1.2.3) which are implied in the reproduction of social (including linguistic) practices.

The corpus collected for this study is well-defined but biased. It includes only certain sectors of the Netherlandic speech community, that is, predominantly the White petty bourgeoisie and the emerging professional class. The 'working classes', on the other hand, which have been of central interest to sociolinguistic research, are largely excluded. As mentioned above, this is not the result of a sampling decision, but reflects the nature of the historical records which not only have been preserved randomly, but reflect the social distribution of literacy which was ethnically and socially skewed. As a result the corpus resembles in its social composition the material collected by Van Oordt (see § 2.1) in the Kaapse Taalargief for the second half of the 18th century (with the exception of the 'new professional class' which is a 19th century phenomenon). The two corpora differ, however, stylistically: while the Taalargief comprises official reports, the corpus collected for this study only includes private material. Both corpora are admittedly of little use for our understanding of the basilectal input (§ 2.2.1, see also the discussion in the epilogue). Describing in detail the more acrolectal forms of CDV can thus only be part of the historical puzzle, but there is no reason to assume that it is an unimportant or uninteresting part.
QUANTIFYING AND CLASSIFYING VARIATION

It is widely acknowledged that the science of taxonomy is one of the most neglected disciplines [...]. The practice of taxonomy has remained intuitive and commonly inarticulate, an art rather than a science.


An Ancient Chinese Classification of Animals
Animals are divided into (a) those that belong to the Emperor, (b) embalmed ones, (c) those that are trained, (d) suckling pigs, (e) mermaids, (f) fabulous ones, (g) stray dogs, (h) those that are included in this classification, (i) those that tremble as if they were mad, (j) innumerable ones, (k) those drawn with a very fine camel's hair brush, (l) others, (m) those that have just broken a flower vase, and (n) those that resemble flies from a distance.

Jorge Louis Borges *Other Inquisitions* (1937-1952)

5.1 Statistics and Variation Studies: More Than a Numbers Game

Linguistic variation, although sometimes categorical, is often a question of frequency. In other words, while there are situations in which one person (or group of people) uses only variant X, and another person (or group of people) only variant Y, often speakers (or groups of speakers) are distinguished by the relative frequency with which they employ the two variants. It appears to be uncontroversial that such situations of variable frequency can be described quantitatively by counting the number of occurrences of each variant for each individual (or group), and then comparing the relative frequencies (usually expressed in percentages) across individuals or groups. In this way numbers are assigned to originally qualitative observations, and as a result a new type of data is produced which is numerical in nature and can be subjected to statistical analysis.

However, to acknowledge that speech is variable does not mean that one attaches any theoretical importance to the observed frequency variation. In other words, quantitative fluctuations might only be secondary phenomena which do not impinge on the categorical nature of language *per se*. Bloomfield (1933: 37-38), for example, has suggested that the 'speech habits of a community' can adequately be described as the sum of all linguistic forms available to the members of the community, regardless of their frequency patterns, while 'genuinely statistical

---

1 Smith (1989: 178-186) dismissed quantitative variation studies loftily as a mere 'numbers game', irrelevant to linguistic theory whose sole concern is the understanding of language as a mental/psychological entity.
observation' is necessary only for the understanding of the process of language change.\(^2\) Chomsky (1965, 1980) has repeatedly argued that frequency variability is part of performance (or E-language) and not competence (or I-language), which alone constitutes the object of linguistic theory. Sankoff (1980b: 77-78) on the other hand, outlining what she calls the 'quantitative paradigm', has argued for the existence of a strong connection between observable performance and underlying competence, and accordingly interprets performance variability as a central aspect of linguistic competence:

My position has been that statistically fluctuating performance data need not be interpreted as reflecting underlying competence that is categorical in nature, and that a paradigm representing competence as containing some probabilistic and nondeterministic components is a better approximation to linguistic reality than one that insists on categoriality and determinacy. Note that this does not imply the nonexistence of categorical rules, but simply the existence as well of probabilistic rules.

Such an assumption does not counter the principle that (socio)linguistic competence is what exists in people's heads; rather, it takes the position that people can internalize rules that are not categorical.

In other words, describing frequencies and establishing probabilities of occurrence, is not 'merely' a descriptive exercise akin to butterfly collecting, but of theoretical importance for our understanding of language, since probabilities of occurrence describe (however crudely) an important aspect of the nature of the individual's competence.

Once the data are reordered in terms of frequencies and proportions, descriptive and inferential statistical procedures can be used to interpret the numerical information. That the discovery of *patterns* of usage, which pertain to the relative frequency of occurrence or co-occurrence of structures, is a central aim of sociolinguistic research was emphasised by Poplack (1993: 258), and to investigate such patterns the tools of descriptive statistics are useful. This includes numerical summaries (such as mean, standard deviation, median, interquartile range etc.), correlational methods, and different types of visual representations; in short all techniques which fall under the heading of 'exploratory data analysis' or EDA (see Tukey 1977, Diaconis 1995).\(^3\) Descriptive statistics are best described as 'tools of discovery' (Anderberg 1973: 4): they allow the researcher to explore unknown data and to generate hypotheses about the data, which

---

\(^2\) This remains true for Lightfoot's generative account of language change. Lightfoot (1991: 162-163) interprets the surface variability observable in the historical sources as a reflection of the changing triggering experiences which ultimately (i.e. once they change in 'some critical fashion') lead to new parameter settings in the minds of individual speakers (and thus to changes in the I-language).

\(^3\) Correlational methods are central to sociolinguistic research as they summarise the strength of the relationship between linguistic variation and social structure (extralinguistic variables). Strong correlations between a linguistic variant X and a social characteristic Y are generally seen as indicating social meaning of the linguistic variant.
can then be tested against additional data. The strength of the evidence can furthermore be assessed with the help of inferential statistics. ¹

Inferential statistics are used to establish the probability that the relations observed in the sample are likely to occur also in the population from which the sample is drawn, and are not due to a sampling error (for a good summary of the underlying mathematical reasoning see Kretzschmar & Schneider 1996: 21-25). The probability is represented by the so-called p-level. The higher the value of p, the less we can believe that the observed relations between variants in the sample are a reliable indicator of the relations between the respective variants in the population. ⁵ Results that yield p < 0.05 are generally considered statistically significant, i.e. in this case the patterns observed in the sample can be generalised to the population with a specifiable degree of confidence (the probability of error is 5%). ⁶ As sociolinguistic samples (especially historical ones) are often convenience samples (see § 4.1.2), it is necessary to adopt a 'pragmatic' view with respect to the interpretation of the results from inferential statistics, i.e. one should carefully consider the conditions under which the data was gathered, and interpret the results of the statistical analysis with these factors in mind (Woods, Fletcher & Hughes 1986: 54-57).

While significance is a binary concept, that is, either a relation between variables is statistically significant according to a pre-defined p-level or not (as Kretzschmar & Schneider 1996: 23, aptly put it 'statistical significance is like pregnancy: either she is or she isn't'), it is generally useful to indicate the actual value of the p-level, as its magnitude might suggest further leads for research.

The notion of significance is [...] different from the probabilities that are used to determine it. It is certainly true that one may call a particular event less probable as the p value approaches 0.0 (a zero percent probability), or more probable as the p value approaches 1.0 (a 100% probability). [...] Whether a measurement has a p<.50 value (a 50/50 probability) or a p<.10 value or a p<.051

---

¹ Early sociolinguistic work in the 1960s has relied almost exclusively on descriptive statistics, and inferential methods were only used from the early 1970s (see the overview given by Davis 1982).

⁵ In more technical terms, inferential statistics are used to prove or disprove what is called the null hypothesis (H₀). The null hypothesis states that the relations observed in the sample do not exist in the population from which the sample is drawn. The alternative hypothesis (H₁) states that the distribution observed in the sample is also true for the population. The strategy of inferential statistics is to accumulate enough evidence to reject the null hypothesis, and a p-level of 0.05 indicates that the probability of erroneously rejecting the null hypothesis is 5%.

⁶ The choice of the p-level depends of course on the nature of the research project, i.e. while a 5% error probability seems acceptable for most studies in the social sciences, it would be rather high when it came to testing the potential harmfulness of a new drug, in such cases p-levels of p < 0.01 or even p < 0.001 are often used (Kretzschmar & Schneider 1996: 22).
value can make a great difference for future testing (either collection of more data or performance of different statistics) even though it is not significant for the test at hand. (Kretzschmar & Schneider 1996: 23)

Inferential statistics can be divided into parametric and non-parametric methods; the former demand a moderate to large sample size and are based on the assumption that the scores are normally distributed. Nonparametric tests, on the other hand, do not rely on the estimation of population parameters, nor do they need large samples. In other words, they are specifically designed for the analysis of low-quality data, that is, small samples drawn from populations about which little is known. Probably the best-known non-parametric technique is the \( \chi^2 \)-test, which measures the significance of the deviation of a given sample distribution from the expected frequency of events \( \chi^2 = \sum (f_{\text{obs}} - f_{\text{exp}})^2 / f_{\text{exp}} \). The test can be used whenever we have a cross-tabulation of two variables (i.e. we can have as many columns as we like as long as we have only two rows; the overall structure of the data must remain binomial).\(^7\)

Although the data analysis presented in chapters six, seven and eight uses inferential statistics \( (\chi^2\text{-test}) \) to assess the reliability of distribution patterns noticeable in the data, the core of the analysis is descriptive, including both uni-, bi- and multi-variate techniques. While uni- and bi-variate techniques are well-established in sociolinguistic research, the question of how several individual variables (and their varying frequencies) are combined into linguistic sets or ‘varieties’ has received comparatively little attention (see Horvath 1985: 153-154, Cheshire 1987: 278). The following sections therefore provide an overview of the multivariate techniques used in the analysis of the corpus data. All these techniques are essentially taxonomic or classificatory, grouping either the cases or the variables on the basis of the numerical information provided in the data set.

5.2 Multivariate Clustering Techniques
5.2.1 Numerical Taxonomy

Classification is not only a basic human conceptual activity (see Lévi-Strauss 1972: 4-8), but also a fundamental process in the practice of science. Observation is inevitably followed by the systematic cataloguing of objects in terms of their properties and relations, and the resulting classifications help to structure the domains of scientific inquiry. More than 30 years ago the biologist G.G. Simpson stated the importance and value of scientific classifications emphatically in his Animal Taxonomy (1961):

---

\(^7\) The application of probability theory in variation studies is not limited to the use of general statistical techniques, but also led to the development of VARBRUL. VARBRUL (see Cedergreen & Sankoff 1974) assigns probabilities to the linguistic and social constraints which affect a variable rule, and can thus be said to re-define the notion of linguistic optionality in terms of statistical probability theory.
Scientists do tolerate uncertainty and frustration, because they must. The one thing they do not and must not tolerate is disorder. The whole aim of theoretical science is to carry to the highest possible and conscious degree the perceptual reduction of chaos [...]. In specific instances it can well be questioned whether the order so achieved is an objective characteristic of the phenomena or is an artificial construct by the scientist. [...]. Nevertheless the most basic postulate of science is that nature itself is orderly [...]. All theoretical science is ordering and if systematics is equated with ordering, then systematics is synonymous with theoretical science. (Quoted in Lévi-Strauss 1972: 9-10)

Taxonomists have frequently based their classifications on their experience with the data and their resulting intuitions about its structure. In the 1960s, however, objections were raised against such 'subjective' methods and a multivariate, strictly empirical approach to classification, relying on explicitly formulated statistical methods rather than intuition, was advocated by the biologists R. Sokal and P. Sneath. In their book *Principles of Numerical Taxonomy* (1963) Sokal and Sneath argued that in order to construct adequate biological classifications one should gather all information available on a set of organisms, code this information numerically, and then use a clustering algorithm to group similar objects together into groups. Sokal & Sneath's book was a major stimulus for the development of automatic clustering techniques, and the increasing availability of high speed computers, which reduced computing time and made it possible to handle large data sets, further encouraged interest in and application of clustering techniques in many different fields. Today cluster analysis, multidimensional scaling and principal components analysis (PCA) are widely used for establishing classifications, not only in botany and zoology, but also in the medical sciences, sociology, engineering and economics.

In quantitative sociolinguistic research clustering techniques have also been shown to be powerful heuristic tools for exploratory data analysis; however their use is by no means as widespread as in other fields. Cluster analysis was used in the Tyneside sociolinguistic survey to classify speakers into groups based on similarities in their linguistic behaviour (Pellowe 1976: 210-212; see also Schneider & Viereck 1984: 40-43). The sociolinguistic survey of multilingual communities which was carried out by LePage *et al.* in Belize and St.Lucia also made use of cluster analysis, however with limited success (McEntegart & LePage 1982: 119-124, LePage & Tabouret-Keller 1985: 80). Bernstein (1993) used cluster analysis in her study of phonological variation in Texas to reduce the original 12 linguistic variables into a smaller number of clusters. Multidimensional scaling and PCA was applied by Berdan (1978) to Macaulay's Glasgow data for the investigation of intra-group variation, and by Cichocki (1988) for the analysis of regional
the data on complement *que*-deletion in Montreal French. PCA was used by Ma & Herasimchuk (1972) in their analysis of speech styles of Puerto Rican bilingual speakers to show the interdependency of phonological variables, as well as by Horvath (1985) in her study of variation in Australian English. Clustering techniques have also been used to group different speech and text genres in terms of their linguistic similarity (Biber 1988, Besnier 1988). Despite the successful application of multivariate clustering techniques in sociolinguistic research, uni- and bi-variate data summaries have generally been dominant. Thus, while statistical concepts such as 'mean', 'standard deviation', 'p-level' and 'significance' certainly belong to the academic jargon of sociolinguists, the same is not true for terms such as 'linkage distance', 'amalgamation rule', 'alienation coefficient' 'eigenvalue' or 'component loading', which are integral to the clustering techniques used in this study. It remains, therefore, necessary that linguists who decide to make use of lesser known techniques introduce them carefully. Pellowe (1976: 203-204) expressed this need for descriptive transparency decisively:

Unclearness in the specification of aims and assumptions, allusiveness in a statement of methods, suppression of exceptions, and many other features of what may be normal presentation methods in other research areas, are particularly disadvantageous for the proper growth of sociolinguistic research, since precisely this information must be the basis for the formation of a critical apparatus by which to evaluate competing approaches.

The following sections offer a conceptual and operational, rather than mathematical, description of the three multivariate techniques used in chapter eight. Furthermore, based on a small linguistic data set (Labov 1969) a step-by-step description of the methods, including all statistical outputs, will be given.

5.2.2 Hierarchical Cluster Analysis

The conceptual and mathematical simplicity of cluster analysis is striking when compared to the complex deductive mathematics underlying other statistical methods. In cluster analysis individual cases (these are usually the speakers or variables in sociolinguistic research) are

---

9 Statistical textbooks for students of linguistics (such as Butler 1985, Woods, Fletcher & Hughes 1986) generally focus on uni- and bi-variate techniques, and offer little guidance to the student who intents to use multivariate procedures (see also the review by Van Hout 1988: 978). The more recent textbook by Rietfeld & Van Hout (1993), however, includes a comprehensive chapter on PCA.

10 Cluster analysis is known under a variety of names, such as 'numerical taxonomy', 'automatic classification', 'betology', 'phenetics' and 'typological analysis' (Kaufman & Rousseau 1990: 3). An accessible description of cluster analysis for the mathematically untrained is given by Aldenderfer & Blashfield (1984), a more technical up-to-date discussion with extensive references by Arabie & Hubert (1995). Romesburg's *Cluster analysis for researchers* (1984) offers detailed information and gives a summary of the mathematics involved.
grouped together in terms of their similarities, which are calculated on the basis of the metric distances between cases. The metric concept of similarity can be illustrated by using a two-dimensional example (following the discussion in Romesburg 1984).

<table>
<thead>
<tr>
<th>variable</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>case</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>30</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 5.1 Two-dimensional data set.

The rows in this small data set stand for cases, i.e. objects whose similarity to each other we want to estimate. The columns stand for variables, i.e. the properties of the cases. We can treat the two variables as coordinates that fix the positions of cases 1-4 in a two-dimensional space (figure 5.1). The distances between the cases 1-4 are indicated by dotted lines and can be interpreted as indices of similarity or dissimilarity. Examining the distances visually, the cases one, four and possibly three can be described as being relatively similar to one another and thus forming a cluster, while the distance of case number two indicates a high degree of dissimilarity to the other three cases.

Figure 5.1 Position of cases 1-4 in the two-dimensional space.
A convenient measure for the distances (i.e. the dissimilarities) between cases 1-4 can be obtained from the Euclidean distance which is computed by using the Pythagorean theorem (for details regarding the mathematical procedure see Romesburg 1984: 13-14, Kaufman & Rousseeuw 1990: 11-12). Briefly, following Pythagoras' theorem \( a = \sqrt{b^2 + c^2} \) the distance \( a \) between two points \( i \) and \( j \) with the coordinates \((x_i, y_i)\) and \((x_j, y_j)\), is given by the length of the hypotenuse of the triangle and takes the following value:

\[
a(i, j) = \sqrt{(x_i-x_j)^2 + (y_i-y_j)^2}.
\]

This expression can be extended to an \( n \)-dimensional space (i.e. situations where we have \( n \) variable values under consideration) as follows: \( d(i, j) = \sqrt{(x_{i1}-x_{j1})^2 + (x_{i2}-x_{j2})^2 + \ldots + (x_{in}-x_{jn})^2} \), where \((x_{ix}, x_{jx})\) represents the variable values for the objects \( i \) and \( j \) (see Everitt 1986). Thus, two variables create a two-dimensional space, three variables a three dimensional space, and so forth.\(^{11}\)

\[\text{Figure 5.2. Illustration of the Euclidean distance formula (adapted from Kaufman & Rousseeuw 1990: 12).}\]

The distances between the four cases are then displayed in a distance or dissimilarity matrix (table 5.2). Such distance matrices are the first step in the development of numerical classifications.

\(^{11}\) Other types of distance measures are available in STATISTICA: squared Euclidean distance, Manhattan (city block) distance, Chebychev distance, power distance. Apart from truly geometrical (i.e. metric) distance measures, correlation coefficients can be used to calculate the similarities between objects on the basis of their correlations (Aldenderfer & Blashfield 1984: 22-24).
Table 5.2 Distance or dissimilarity matrix for cases 1-4 (Euclidean distances).

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>22.4</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>11.2</td>
<td>18.0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5.0</td>
<td>20.6</td>
<td>14.1</td>
<td>0</td>
</tr>
</tbody>
</table>

After the distances between the objects have been established, cases are linked together into clusters at certain intervals. That is, a clustering algorithm mechanically searches the distance matrix for the lowest distance value between two cases, and then combines these two cases into a cluster. In the example data set, this means that cases 1 and 4 would be linked. Now the distances between the remaining cases and the cluster \( \{1,4\} \) will be searched by the algorithm, and again the two items with the smallest distance value are merged (for more details see Everitt 1986). Hierarchical clustering techniques are thus agglomerative, and successively combine cases into larger and larger clusters by grouping at any step the two cases (or groups of cases) that are closest to each other together into a new cluster.

Two main types of cluster linkage (or merger) rules can be distinguished: linkage of nearest neighbour (single linkage) and linkage of furthest neighbour (complete linkage). Linkage of nearest neighbour means that an object (or a cluster of objects) is joined to an existing group based on a high level of similarity with any member of the existing group; that is, only a single link is required for a case (or groups of cases, i.e. a cluster) to merge with an existent cluster (figure 5.3). A major drawback of single linkage is that it has a tendency to form elongated clusters (so-called 'chains'), in which some members are very far from each other.

![Figure 5.3 Single linkage (adapted from Kaufman & Rousseuw 1990: 47).](image)

Complete linkage (i.e. linkage according to the furthest neighbour) is the opposite of single linkage: a case will only be merged with an existing cluster if this case shows a minimum level of similarity to the furthest member of that cluster (figure 5.4). As a result the data set will be patterned into relatively compact clusters containing highly similar cases (Aldenderfer & Blashfield 1984: 40). Complete linkage is usually recommended in cases where the objects actually form naturally distinct groups; it should be used with care in cases where no obvious
patterning is to be expected (STATISTICA 1995 III: 3171).

As a compromise between single and complete linkage, average linkage was developed (Sokal & Sneath 1963: 182-185). Under this algorithm an average of the similarity between two clusters is computed (by averaging the length of the distances between all objects of cluster A and all objects of cluster B), and the clusters are joined only if a minimum level of similarity can be established on the basis of this average value. Average linkage has been shown to perform well on many types of data and is widely used (STATISTICA 1995 III: 3171; Romesburg 1984: 126). Apart from average linkage, Ward's method is a frequently used linkage rule (Ward 1963, Romesburg 1984: 129, Jobson 1992: 524). Ward's method was designed to minimise the within-cluster-variability by using an analysis of variance approach to evaluate the distances between clusters. In several studies Ward's method has been shown to outperform other clustering methods, including average linkage (Aldenderfer & Blashfield 1984: 61; for a detailed description of Ward's method see Romesburg 1984: 129-135).

A hierarchical tree diagram (also called dendrogram) is the graphical output of cluster analysis. The diagram represents the hierarchical relationships that exist between the different cases and clusters, and illustrates the series of steps which led to the final classification, that is, each step at which a pair was merged is represented as a node in the tree. Unlike the evolutionary tree diagrams used in biology and linguistics, the clustering tree only represents a classification of cases based on affinity, but not necessarily phylogeny (descent). The major steps of cluster analysis can be summarised as follows:

---

12 In biological taxonomy, the so-called 'transformed cladists' reject the postulate that classification and phylogeny should be isomorphic, 'but accept as axiomatic the assumption that the natural order of organisms is a divergent hierarchy' (Panchen 1992: 60).
The decision for the use of any of the distance measures or linkage rules lies entirely with the researcher, who therefore needs to be aware of the strengths and weaknesses associated with the different methods.

The interpretation and especially the validation of the clustering results are the most difficult steps in cluster analysis:

An inherent problem in the use of a clustering algorithm in practice is the difficulty of validating the resulting data partition. This is a particularly serious issue since virtually any clustering algorithm will produce partitions for any data set, even random noise data which contains no cluster structure. Thus, an applied researcher is often left in a quandary as to whether the obtained clustering of a real life data set actually represents a significant cluster structure or an arbitrary partition of random data. (Milligan 1981, quoted in Jobson 1992: 563/564)

In other words, in cluster analysis we cannot test the null-hypothesis (i.e. the hypothesis that there is no structure in the data), and it is thus difficult to determine whether the clusters identified by the algorithm reflect 'true' or 'natural' typologies, or whether they constitute 'administrative clusters' which simply present more or less convenient groupings of the data set. The validity of a cluster structure can be assessed by considering the following criteria (for summaries of validation procedures see also Romesburg 1984: 256-259, Aldenderfer & Blashfield 1984: 62-74, Jobson 1992: 563-568):

- External criteria can be used to validate a given cluster solution. Validation through the use of external criteria is particularly feasible in sociolinguistic research, where clusters are usually established on the basis of the linguistic variables alone, and the results can then be checked against extra-linguistic information (i.e. whether the clusters are also well-formed in terms of social group membership). In biological taxonomy this is usually discussed under the category of 'predictivity', and a high degree of predictivity (understood as 'the degree to which

---

13 The use of terms such as 'natural clusters' does not imply a realist ontology. The meaning of 'natural' in this context is very different from the Aristotelian one and is based on criteria such as separateness, density and predictive value (see Panchen 1992: 148).
a specific classification agrees with characters not used in the formulation of that classification, Fitch 1979) is seen as an indicator for naturalness (Panchen 1992: 149).

- Internal criteria can also be used to validate the cluster solution and assess the naturalness of the identified clusters. Natural clusters are assumed to satisfy the properties of external isolation (i.e. the clusters are clearly separated from another, the criterion of separation) and internal cohesion (i.e. the cases in a cluster are very similar, the criterion of density; Jobson 1992: 519). The internal structure of the cluster solution can be investigated with the help of k-means clustering (also called 'iterative partitioning method') as well as by inspecting the visual representation of the distance matrix through multidimensional scaling (see § 5.2.3). The input for k-means clustering is a specified number of clusters \( k \) and based on this input the program will compute exactly \( k \) clusters that are as distinct as possible. Based on the descriptive output of k-means clustering, the internal structure of each cluster and the distances between the clusters can be described.

- The replicability of a cluster solution based on split samples of the original data set is an important criterion for the evaluation of a cluster solution. Sokal (1986: 435) distinguishes between 'character stability' (stability of the classification despite the addition or subtraction of variables) and 'case stability' (the classification is robust even if cases are randomly added or subtracted). If the solution cannot be replicated this is reason enough to reject the solution entirely. However, successful replication alone does not guarantee the validity of a solution (Aldenderfer & Blashfield 1984: 65).

- A comparison of cluster solutions obtained by different clustering algorithms, including k-means clustering, and different multivariate techniques (such as principal components analysis and multidimensional scaling), can also be used to validate the cluster solution. In this case validity would be assumed if the results obtained from the different techniques agree with each other. The use of especially k-means clustering for the validation of a cluster solution obtained from the hierarchical analysis also helps to compensate for a general shortcoming of hierarchical methods, which suffer from the defect that they can never repair what was done in a clustering step, i.e. once two objects have been joined into a cluster, they cannot be separated (Kaufman & Rousseeuw 1990: 44-45, Woods, Fletcher & Hughes 1986: 261). K-means clustering, on the other hand, makes several passes through the data and can therefore modify a poor initial partition (Aldenderfer & Blashfield 1984: 46).\(^{14}\)

\(^{14}\) However, it should be noted that partitioning methods have not been studied in as much detail as hierarchical methods, and there is much more insecurity about how one should interpret the results
To illustrate in a little more detail how cluster analysis works, Labov's data on copula deletion in AAVE can serve as an example.

*The Copula in AAVE (Labov 1969)*: The term 'African American Vernacular English' (AAVE) refers to a variety of English learned by most African Americans as a first language, and which is used by the majority of Black adults in relaxed settings when conversing with friends and family (Labov & Harris 1986: 4). AAVE differs considerably from other varieties of American English with regard to grammar, pronunciation and lexicon. A well-known linguistic feature is the optional deletion of the copula in a number of syntactic environments. Labov has shown that variation in the copula is the result of a series of grammatical and phonological rules that are parallel to those of colloquial contraction in standard English (Labov 1969:722ff.)\(^\text{15}\). No AAVE-speaker deletes the copula always - variation of full, contracted, and deleted forms is typical. An important variable constraint on deletion in AAVE (as well as on contraction in Standard English) is the nature of the subject, i.e. whether the subject is a pronoun or some other type of noun phrase. The data set given by Labov lists the percentages for full, contracted and deleted forms of *is* for six groups of informants. The first four groups are different pre-adolescent peer groups (with members aged between 10 and 18 years), 'Adults' refers to a sample of working-class adults in the Cobra and Jet territory, 'Inwood' to a contrasting sample of white working-class speakers from the Inwood neighbourhood of upper Manhattan (table 5.3).

<table>
<thead>
<tr>
<th></th>
<th>NP/full</th>
<th>NP/con.</th>
<th>NP/del.</th>
<th>P/full</th>
<th>P/con.</th>
<th>P/del</th>
</tr>
</thead>
<tbody>
<tr>
<td>T-Birds</td>
<td>44</td>
<td>15</td>
<td>42</td>
<td>7</td>
<td>33</td>
<td>60</td>
</tr>
<tr>
<td>Cobras</td>
<td>45</td>
<td>19</td>
<td>36</td>
<td>0</td>
<td>23</td>
<td>77</td>
</tr>
<tr>
<td>Jets</td>
<td>54</td>
<td>19</td>
<td>27</td>
<td>0</td>
<td>42</td>
<td>58</td>
</tr>
<tr>
<td>Oscar Bros</td>
<td>51</td>
<td>23</td>
<td>26</td>
<td>4</td>
<td>33</td>
<td>64</td>
</tr>
<tr>
<td>Adults</td>
<td>61</td>
<td>26</td>
<td>14</td>
<td>1</td>
<td>72</td>
<td>27</td>
</tr>
<tr>
<td>Inwood</td>
<td>41</td>
<td>59</td>
<td>0</td>
<td>1</td>
<td>99</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 5.3 Percentage of full, contracted, and deleted form of *is* with noun phrase (NP) subject or pronoun (P) subject (group style only); due to rounding some percentages do not add up to 100 (Labov 1969: 730).

(Aldenderfer & Blashfield 1984: 45). In STATISTICA it is possible to control the way in which the initial cluster centres are computed. The following options are available: 'maximize between-cluster distances', 'sort distances and take observations at constant intervals' and 'choose the first N (number of clusters) observations' (for a discussion of the different methods see STATISTICA III: 3187-3188). For the analyses in this study the option 'maximize between cluster distances' was used since it generally yields well-defined and maximally separated clusters.

\(^\text{15}\) Labov's interpretation has since been modified by Baugh (1980) who showed that unlike standard English, AAVE favours zero copula in adjective environments.
Cluster analysis can be performed on the raw or the standardised data matrix. Standardisation is necessary only in cases where different units of measurement are used; however, if the distribution and magnitude of the variables is believed to be significant, standardisation is not advisable. In this case the raw data matrix was used as input into the program as the values of the variables are expressed in percentages and are thus comparable.

Concerning the choice of a distance measure and a linkage rule, it was mentioned above that it is common practice to use different distance measures and linkage rules for exploratory data analysis, and to carefully compare the resulting classifications. Three different linkage rules were used for data exploration: single linkage, average linkage and Ward's method. The distance measure employed is the Euclidean distance (figure 5.5, 5.6 and 5.7).

The map given by the tree shows a hierarchical ordering of similarities that begins at the bottom of the tree where each object or case is separate, in its own cluster and similar only to itself. As one moves upwards individual cases are merged into clusters on the basis of their similarity.

![Tree diagram for single linkage (Euclidean distance).](image)

---

16 In STATISTICA standardisation is computed as follows: Std. Score = (Raw Score - Mean)/ Std. Deviation (STATISTICA 1995 I: 1369).

17 Other distance measures (Manhattan and Chebychev) were also used for data exploration; however, the resulting classifications differed only marginally from the ones established on the basis of Euclidean distance.
A comparison of the three diagrams suggests that we either have two clusters (Jets, T-Birds, OscarBros, Cobras vs. Adults, Inwood) or three clusters (Inwood vs. Adults vs. Jets, T-Birds, OscarBros, Cobras). Furthermore, all three diagrams represent OscarBros and Jets as forming a subgroup within the cluster 'pre-adolescent peer-groups'. The fact that the pre-adolescent peer groups are clustered together was to be expected from our knowledge about the social context of language use in American cities. A second (expected) result is that 'Inwood' is clearly separated from the adolescent peer group cluster. The position of black working class
adults, however, is somewhat ambiguous - do they form a linguistic cluster of their own or should they be grouped together with Inwood?

The interpretation of the hierarchical tree diagram is not always as straightforward as this example might suggest. The more cases involved in the analysis the more complicated and detailed the diagram becomes. The nested structure of the diagram suggests that many different groups exist in the data and the analyst has to decide where to 'cut' the tree in order to obtain the optimal number of clusters (Aldenderfer & Blashfield 1984: 53). Frequently the use of external criteria as well as expert intuition can help with this decision (Anderberg 1973: 176). A more formal approach is the interpretation of the so-called graph of amalgamation (linkage) schedule, in which the successive clustering steps are plotted against the linkage distances at which the clusters were formed (figure 5.8). The graph can be useful for the decision about a cut-off point for the tree diagram.

![Amalgamation Graph](image)

Figure 5.8 Amalgamation graph (Euclidean distance, Ward's method).

The amalgamation graph is read from right to left. A jump in the graph indicates that two relatively dissimilar clusters have been merged, a flattening of the graph implies that after this point clusters were formed more or less at the same linkage distance. The number prior to the point where the graph levels out into a plateau is therefore the most probable solution. In this example a three cluster solution is indicated by the amalgamation graph, i.e. Inwood vs. Adults vs. Jets, OscarBros, Cobras, T-Birds. That this solution is plausible in terms of external criteria does not need detailed explanation. But what can be said about the internal structure of the clusters, how similar or dissimilar are their objects and how distinct are the clusters from each other? In other words, do they fulfill the requirements for so-called natural clusters?
To investigate the internal structure of the three clusters k-means clustering with k=3 can be used. The clusters identified by the partitioning method are identical with the ones obtained from the hierarchical analysis: Inwood vs. Adults vs. pre-adolescent peer groups. The first two clusters are by definition homogeneous as they only contain one member. The structure of the cluster 'pre-adolescent peer groups' shows that the Cobras, which were also singled out in two of the dendrograms, are furthest removed from the cluster centre.

<table>
<thead>
<tr>
<th>Distance from cluster centre</th>
<th>T-Birds</th>
<th>Cobras</th>
<th>Jets</th>
<th>OscarBros</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5.2</td>
<td>6.8</td>
<td>5.8</td>
<td>3.4</td>
</tr>
</tbody>
</table>

Table 5.4 Internal structure of cluster II (pre-adolescent peer groups).

The distances between the clusters indicate that the cluster 'peer group' is separated clearly from 'Inwood', while 'Adults' is located somewhere between 'Inwood' and 'peer groups' (table 5.5).

<table>
<thead>
<tr>
<th></th>
<th>Inwood</th>
<th>Adults</th>
<th>peer groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inwood</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adults</td>
<td>22.9</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>peer groups</td>
<td>43.4</td>
<td>24.2</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Table 5.5 Euclidean distances between clusters.

5.2.3 Multidimensional Scaling

Unlike cluster analysis, multidimensional scaling is not designed to link elements together into clusters; it simply represents the similarity relationships between the cases (based on the values of the distance matrix) in an n-dimensional space by arranging objects in such a way that the original rank order of distances is maintained.\(^{18}\) For example, multidimensional scaling could be used to create a two-dimensional map based only on information about the road distances between a given number of cities or other geographical locations. To achieve this, the program moves the cases in the space defined by the requested number of dimensions (i.e. number of coordinate axes in the \(n\)-dimensional space), and checks how well the original rank order of the

\(^{18}\) There are two different types of multidimensional scaling: classical metric and non-metric scaling. While the aim of non-metric scaling is to reproduce the rank-order of the original distances, classical metric scaling reproduces the actual distances. In this study only non-metric scaling (Guttman Lingoes) was used.
distances between the cases are reproduced by the configuration. On the first iteration, all objects are ordered along a single line, then the reproduced distances along the line are compared with the original rank order of the distances. Typically a large measure of error is found. On the following iterations the cases are moved about until the differences between original and reproduced distances are close to zero (for a detailed description of the iteration procedure see Schiffman, Reynolds & Young 1981: 7-10). To measure what is called the 'goodness of fit' of a configuration (i.e. how well a particular configuration reproduces the original rank-order of the distances), two different indices are used: the stress and the alienation coefficient (for details of the computation see STATISTICA III 1995: 3237, 3255). The smaller the value for both coefficients, the better the fit of the reproduced configuration. The 'goodness of fit' of a particular solution can also be assessed with the help of the so-called Shepard-diagram (named after Roger Shepard's seminal paper of 1962) in which the reproduced distances are plotted against the original distances.

Multidimensional scaling requires as input a distance matrix, which can be obtained from the cluster analysis module. Next, the analyst has to decide on the number of dimensions to be computed. At this point the decision is purely intuitive, but once the dimensions have been established their adequacy can be investigated. Generally one would start with as few dimensions as possible and only if a two-dimensional solution cannot be shown to be adequate should more dimensions be included. If no more than three dimensions are specified, a two or three-dimensional scatterplot can be used to represent the configuration of cases. With help from the results of cluster analysis or simply through visual inspection of the graph, groups of cases can then be identified in the scatterplot. Multidimensional scaling thus comprises the following steps:

**INPUT**
Distance matrix (usually Euclidean).

**STEP 1**
Specification of $n$ dimensions and computation of multidimensional scaling.

**STEP 2**
Review 'goodness of fit' of the configuration (Shepard's diagram, stress and alienation coefficient)

**OUTPUT**
Scatterplot of the configuration.

The interpretation of the results includes the visual inspection of the scatterplot, as well as the interpretation of the actual dimensions established by the program. To interpret the dimensions one looks at the properties of the cases at each end of the axis to determine whether there is some characteristic that changes clearly along the dimension. However, the dimensions are not always interpretable (Schiffman, Reynolds & Young 1981: 13). Replication of the analysis based on split-samples and correlation with external criteria can used for the validation
of the result.

Example (Labov 1969): The dissimilarity matrix (Euclidean distances) obtained from the cluster analysis module was used as input for multidimensional scaling. Two dimensions were specified and eight iterations were computed. In the final configuration stress and alienation coefficient were zero, which suggests a perfect representation of the original data. The Shepard diagram also indicates a flawless re-scaling of the distance matrix (figure 5.10). The steep function depicted in the Shepard diagram represents the monotone transformation of the data, and each data point represents a combination of the original data dissimilarity (plotted on the X-axis) and the reproduced distances (plotted on the Y-axis). A close fit of the steep function to the data points indicates a very good fit between original and reproduced configuration (for details regarding the computation, see STATISTICA 1995 III: 3255).

![Figure 5.9 Shepard's diagram.](image)

The clusters established with the help of the cluster analysis algorithms can easily be located in the scatterplot which depicts the final two-dimensional configuration (figure 5.11). Each point in the plot corresponds to a case (here group of speakers). Dimension I distinguishes the cases with regard to the variables involving contraction and deletion of the copula, i.e. cases on the left show high scores for deletion and low scores for contraction; the opposite is true for cases on the right. Dimension II separates 'Adult' from the other cases, and considering the original data this can be interpreted as contrasting high and low scores on the variable NP/full.
Based on the pictorial representation in the scatterplot, the clusters can be described in terms of density, variance, shape and separation (Aldenderfer & Blashfield 1984: 34). A thick swarm of data points in a space constitutes a cluster of high density, while data points are widely separated in a cluster of low density. Variance refers to the dispersion of the data points from the centre of the cluster. In 'tight' clusters all data points are near the centroid, while the data points are dispersed from the centre in 'loose' clusters. Finally, shape is simply the arrangement of points in the n-dimensional space, and separation is the degree to which clusters are separated from one another.

5.2.4 Principal Components Analysis (PCA)\textsuperscript{19}

Unlike cluster analysis where the researcher can decide on the setting of many parameters (distance measures and linkage rules), most of the framing decisions in PCA have already been made by statisticians and the method can, therefore, not easily be tailored to meet the specific needs of a given data set or research question. The central idea of PCA is to reduce the dimensionality of the data set, which is believed to consist of a large number of interrelated variables. By transforming a set of related variables into a smaller set of uncorrelated

\textsuperscript{19} Jackson's \textit{User's guide to principal components analysis} (1991) gives a detailed description of the procedure, including the mathematical foundations. A more conceptual description is Stevens (1992). PCA differs from the related method factor analysis, in that no assumptions are made about the distribution of the data, and the results are purely descriptive. Factor analysis, on the other hand, assumes normal distribution and has an inferential aspect (for a detailed comparison of factor analysis and PCA see Rietveld & Van Hout 1992: 251ff.).
components, the number of variables is reduced and the structure of the relationship between the variables can be discovered. To understand how variables are combined into components, it is easiest to look again at a two-dimensional example. The relationship between the deletion of the copula in AAVE after pronoun subjects and noun phrases can be summarised in a scatterplot with [NP/del.] on the horizontal axis and [P/del.] on the vertical axis (figure 5.11).

![Scatterplot](image.png)

Figure 5.11 Correlation between [NP/del.] and [P/del.] with a least squares fit (regression line).

The graph indicates that the two variables are highly and positively correlated (correlation coefficient $r=0.9$). A regression line (eigenvector), which captures most of the variance displayed by the two variables, has been fitted to summarise their linear relationship. In order to account for the remaining variability a second line can be drawn through the data, then a third one, and so forth. In this way consecutive eigenvectors or components are extracted. If the two-variable plane is extended to a multidimensional space, the computations become more involved, but the principle remains the same. For each eigenvector a so-called eigenvalue is computed, which indicates the amount of variance extracted by the component.\textsuperscript{20} The eigenvector with the highest eigenvalue becomes the first principal component and accounts for the greatest percentage of variance in the data. Each consecutive component is defined to capture the maximum amount of the variability not captured by the preceding component. Generally a small number of these components will account for most of the variance in the data set. However, if the objective of PCA is to replace $n$ variables by a smaller number, $m$, of principal components, it is important

\textsuperscript{20} For details concerning the computation of eigenvalues, see STATISTICA (1994 III: 3231) and Jackson (1991: 7-10).
to know how small \( m \) can be taken without serious information loss. Various rules have been proposed for guiding the decision of how many principal components to retain:

- The most commonly used criterion is probably Kaiser's rule. Kaiser (1960) proposed to retain only components with eigenvalues greater than 1 as components with eigenvalues lower than 1 extract less variance than one of the original variables (STATISTICA 1995 III: 3204). Several studies have shown that the criterion is fairly accurate, especially when the number of variables is low (10-15) or moderate (<30; Stevens 1992: 378). Blind use of Kaiser's rule can, however, lead to the preservation of principal components which have no practical significance (in terms of the percentage of variance accounted for; Stevens 1992: 378).

- A graphical method, the scree test, has been proposed by Cattell (1966). To construct the graph, the eigenvalues of the principal components are plotted against their ordinal numbers (whether it was the first principal component, the second etc.). Typically the graph shows a sharp drop at one stage and then levels off. The interpretation of the scree graph is analogous to the interpretation of the amalgamation graph in cluster analysis: first the point beyond which the graph levels out is established, then the first point in the already levelled graph is taken to be the last component to be retained (see the example below). Components identified by this method generally account for most of the variance in the data; however, factors which account only for small amounts of variance, but are nevertheless important for our understanding of the correlation structure, might be lost (Stevens 1992: 378).

- A rather straightforward criterion for the retention of principal components is to define a percentage of variation (75, 80 or 90%), and to retain as many components as are necessary to account for the specified amount of variance.

All three rules are \textit{ad hoc} and their justification lies primarily in the fact that they are plausible and work in practice (Jolliffe 1986: 93). Often the rules are used in combination, i.e. the analyst compares the results from the different rules and then decides on the number of components to retain.

The interpretation of the components constitutes the next step in the analysis and is based on the component loadings. The loadings indicate the strength of the correlation between the original variables and the components, and allow the analyst to trace the structure of the relationships between variables. However, often the component loadings are not easily

\[21\] Many statistical programs use this criterion automatically. STATISTICA, however, gives the option of retaining factors with an eigenvalue <1.
interpretable; in this case the analyst can employ certain techniques to simplify the structure of the component loadings. The most common technique is to rotate the axes of the variable space (i.e. the eigenvectors) in different directions without changing the relative locations of the data points. Two types of rotation will be important in the context of the example discussed below: quartimax and varimax rotations. Quartimax rotations are performed so that each original variable loads mainly on one component, while varimax rotations lead to components loading very high on a small number of variables and very low on other variables (Stevens 1992: 380, Rietveld & Van Hout 1992: 279-280). It should, however, be noted that components are artificial variables and are not necessarily interpretable in any meaningful way (Stevens 1992: 378).

So far PCA has been described as a tool for the grouping of variables, i.e. it enables the analyst to understand the correlations or interrelationships between the individual variables. However, PCA can also facilitate the classification of cases into groups. Given a high correlation between two or more variables, it can be concluded that they are quite redundant, in other words, if it is possible to summarise 30 variables into 3 components, the classification of cases can be based on the scores they show for these components. The component scores can be used as input data for cluster analysis or simply be represented on a scatterplot (as done by Horvath 1985).  

The most important steps of PCA can be summarised as follows:

<table>
<thead>
<tr>
<th>INPUT</th>
<th>Raw data matrix (or correlation matrix).</th>
</tr>
</thead>
<tbody>
<tr>
<td>STEP 1</td>
<td>Computation of PCA.</td>
</tr>
<tr>
<td>STEP 2</td>
<td>Review eigenvalues, scree test and % of variants accounted for and decide on how many components to retain.</td>
</tr>
<tr>
<td>STEP 3</td>
<td>Review component loadings.</td>
</tr>
<tr>
<td>STEP 4</td>
<td>Rotation of factors (optional).</td>
</tr>
<tr>
<td>STEP 5</td>
<td>Interpretation of factors as combinations of variables.</td>
</tr>
</tbody>
</table>

Before turning to the AAVE example two problems concerning the use of PCA are worth

---

22 For a general discussion of the graphical representation of data obtained from PCA see Jolliffe (1986: 64-71). Note that principal components analysis tends to blur the relationship between clusters: widely separated clusters are normally maintained, while the distances between groups that are not widely separated are reduced (Aldenderfer & Blashfield 1984: 21).
mentioning: the type of data matrix and sample size. Principal components analysis can be performed on a raw data matrix or a correlation matrix. Stevens (1992: 376) argued that it is preferable to use the raw data matrix as input into the program, while Jolliffe (1986: 16-19) supports the use of correlation matrices (see also Jackson 1991: 64-66). The problem is similar to the one regarding data standardisation, and no hard and fast rule can be formulated. Considerable differences between the components extracted from the correlation matrix and the raw data matrix can be expected in cases where large differences are found in the standard derivations of the variables (Jolliffe 1986: 33). In such cases it is advisable to use both types of matrices for data exploration and to compare the results.

A second problem concerns the sample size required for reliable components. Generally sample size poses no problem in cluster analysis and multidimensional scaling; however, the situation with regard to PCA is not clear. Several popular rules suggest that the ratio of cases to variables is the crucial criterion (Horvath 1985: 56). There should be more speakers than variables, and the recommended speaker/variable ratio varies from two cases per variable to 20 cases per variable (Stevens 1992: 384). However, a recent Monte Carlo study by Guadagnoli & Velicer (1988, reported in Stevens 1992: 384) indicates that the magnitude of factor loadings is more important than case/variable ratio considerations. Based on Guadagnoli & Velicer's study Stevens (1992: 384) established the following guidelines:

1. Components with four or more loadings above .60 in absolute value are reliable, regardless of sample size.
2. Components with about 10 or more low (.40) loadings are reliable as long as sample size is greater than about 150.
3. Components with only a few low loadings should not be interpreted unless sample size is at least 300.

Furthermore, any component with at least 3 loadings over 0.80 will be reliable (Stevens 1992: 384).23 Regarding the validation of the results, the techniques mentioned with respect to cluster analysis, especially validation through external criteria and replication, are also used in PCA.

---

23 It should be noted in this context that the components extracted in Horvath’s analysis of Australian English do not fulfill these criteria, and their loadings lie typically around 0.30 (Horvath 1985: 57, 72, 76, 84). Factors with loadings > 0.30 were seen as reliable for a long time (Stevens 1992: 383); however, based on the results from Guadagnoli & Velicer’s Monte Carlo study and the table of critical values given by Stevens (ibid.), the sample size in Horvath’s study should have been at least 300 to allow for the significance of component loadings around 0.30. This technical objection should, however, not distract from the fact that Horvath’s results make sociolinguistic sense, which shows that technical guidelines must not be overvalued.
Example (Labov 1969): It has been mentioned above that although the optimal performance of PCA is partially dependent on sample size, the absolute magnitude of the component loadings can outweigh the disadvantage of a small sample. The data set given by Labov for copula deletion in AAVE has a case/variable ratio of only 1:1, but since it shows very high and therefore (probably) reliable component loadings it will be used to illustrate PCA. The analysis was performed on the basis of the raw data matrix and it was specified that only components above an eigenvalue of 0.5 should be extracted (table 5.6).

<table>
<thead>
<tr>
<th></th>
<th>Eigenvalue</th>
<th>% of variance accounted for</th>
</tr>
</thead>
<tbody>
<tr>
<td>Component I</td>
<td>3.90</td>
<td>65</td>
</tr>
<tr>
<td>Component II</td>
<td>1.32</td>
<td>22</td>
</tr>
<tr>
<td>Component III</td>
<td>0.71</td>
<td>12</td>
</tr>
</tbody>
</table>

Table 5.6 Principal components (Eigenvalue > 0.5).

While Kaiser's rule states that only components with an Eigenvalue > 1 should be maintained, the threshold was lowered in this case because the structure of the scree plot suggested that to retain three components would be appropriate (figure 5.12).

![Scree plot](image)

Figure 5.12 Scree plot.

The examination and interpretation of the (unrotated) component loadings constitutes the next step in the analysis (table 5.7).
Component I shows a mixture of high positive and high negative loadings and can be described as a 'bipolar component', contrasting contraction and deletion of the copula. That is, a pattern of high scores on the deletion variables and low scores on contraction variables accounts for most of the variability in the data, thus separating the peer-groups from both Adults and Inwood. The second component shows only one very high (negative) and one moderately high (positive) loading, both on the two variables which were not accounted for in component I. The loadings suggest that very low scores on variable [NP/full] correlate with relatively high scores on variable [P/full]. Component III shows only one moderate (negative) loading (on variable [P/full]). The pattern can be clarified with the help of factor rotation (table 5.8, 5.9).

<table>
<thead>
<tr>
<th></th>
<th>component I</th>
<th>component II</th>
<th>component III</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP/full</td>
<td>0.04</td>
<td>-0.99</td>
<td>-0.14</td>
</tr>
<tr>
<td>P/full</td>
<td>0.18</td>
<td>0.14</td>
<td>0.97</td>
</tr>
<tr>
<td>NP/cont.</td>
<td>-0.90</td>
<td>0.38</td>
<td>-0.19</td>
</tr>
<tr>
<td>P/cont.</td>
<td>-0.99</td>
<td>-0.02</td>
<td>-0.11</td>
</tr>
<tr>
<td>NP/del.</td>
<td>0.94</td>
<td>0.07</td>
<td>0.30</td>
</tr>
<tr>
<td>P/del.</td>
<td>0.99</td>
<td>0.03</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Table 5.8 Component loadings: varimax rotation.

<table>
<thead>
<tr>
<th></th>
<th>component I</th>
<th>component II</th>
<th>component III</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP/full</td>
<td>0.06</td>
<td>-0.99</td>
<td>0.14</td>
</tr>
<tr>
<td>P/full</td>
<td>0.30</td>
<td>0.15</td>
<td>-0.94</td>
</tr>
<tr>
<td>NP/cont.</td>
<td>-0.93</td>
<td>0.34</td>
<td>0.11</td>
</tr>
<tr>
<td>P/cont.</td>
<td>-0.99</td>
<td>-0.05</td>
<td>0.01</td>
</tr>
<tr>
<td>NP/del.</td>
<td>0.96</td>
<td>0.10</td>
<td>-0.21</td>
</tr>
<tr>
<td>P/del.</td>
<td>0.99</td>
<td>0.04</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Table 5.9 Component loadings: quatrimax rotation.
The bipolar structure of component I was not affected by the rotations. The structure of components II and III, however, is now more pronounced: component II loads high on variable [NP/full], while component III loads high on variable [P/full]. However, a component defined by only one loading is in fact not much of a component; in this case the actual variable can simply be substituted. Note also that the total percentage of variance accounted for by the (unrotated) third component is only 12%, while component I and II together account for roughly 87% of the variance. Thus in disagreement with the structure of the scree plot, only two components will be retained. Finally, PCA can be used to cluster cases on the basis of their component scores (table 5.10).

<table>
<thead>
<tr>
<th></th>
<th>component I</th>
<th>component II</th>
</tr>
</thead>
<tbody>
<tr>
<td>T-Birds</td>
<td>0.90</td>
<td>1.27</td>
</tr>
<tr>
<td>Cobras</td>
<td>0.71</td>
<td>-0.60</td>
</tr>
<tr>
<td>Jets</td>
<td>0.25</td>
<td>-0.91</td>
</tr>
<tr>
<td>OscarBros</td>
<td>0.45</td>
<td>0.09</td>
</tr>
<tr>
<td>Adults</td>
<td>-0.54</td>
<td>-1.31</td>
</tr>
<tr>
<td>Inwood</td>
<td>-1.77</td>
<td>0.91</td>
</tr>
</tbody>
</table>

Table 5.10 Component scores (based on unrotated components).

Component I exhibits positive scores for the preadolescent peer groups, i.e. they show high frequencies of copula deletion together with low frequencies of contraction, a pattern most visible in the linguistic performance of the T-Birds. Adults and particularly the Inwood control group show the opposite behaviour (negative loadings on component I). It has been mentioned above that high loadings on component II (unrotated) indicate low scores on variable [NP/full] together with relatively high scores on variable [P/full]. Only T-Birds (and to a lesser degree Inwood) show a pattern of low scores on variable [NP/full] and (comparatively) high scores on variable [P/full]. Jets and Adults, on the other hand, demonstrate the opposite pattern, i.e. relatively high scores on variable (NP/full) and low scores on variable [P/full].

The component scores can be used as the input for cluster analysis, or simply be plotted on a two-dimensional scatterplot to graphically represent the relative location of the cases. Figure 5.13 shows the scatterplot of the first two principal components (thus accounting for around 90% of variance in the data). Note that the internal structure of the cluster 'pre-adolescent peer group' has changed slightly when compared with the results from cluster analysis and multidimensional scaling, i.e. the cluster is less compact and slightly elongated.
5.3 Summary

The multivariate techniques discussed in this chapter are best characterised as classificatory or taxonomic, grouping cases and/or variables on the basis of either similarity (cluster analysis and multidimensional scaling) or correlation (PCA). The importance of multiple similarity relations for the identification of linguistic sets or varieties was emphasised by Lieb (1993: 59-61), and Hudson (1996: 51, 68) has argued that language varieties can be conceptualised in terms of co-occurrence (i.e. co-variation or positive correlation) of individual variables (on the conceptualisation of linguistic varieties in terms of co-occurrence see also Labov 1971: 462, Rickford 1980: 169, Horvath 1985: 154, Schönfeld 1985: 208, Berutto 1987: 264). Techniques such as cluster analysis, multidimensional scaling (describing metric similarities) and PCA (grouping variables on the basis of their co-occurrence) translate such conceptual generalisations into scientific understanding, and are thus promising tools for exploring the structure of linguistic variation.

An important advantage of these techniques is that no a priori hypothesis regarding the existence of sociolinguistic groups is needed, since groups are established purely on the basis of linguistic similarities between speakers. The resulting classifications can then be compared with the social group membership of cluster members, as well as with popular taxonomies of linguistic variation (see § 8.1.1). This approach thus differs from the classical Labovian tradition where the population is divided according to pre-established socioeconomic categories, and it is taken for granted that these social groups share a common linguistic behaviour (see § 1.3; for a critique of the Labovian approach see Bailey 1980: 242, Horvath 1985: 63, Milroy 1988: 580-
Although considering similarities between speakers on a number of different variables will contribute to our understanding of the structure of large-scale linguistic varieties, such methods can cloud the importance of individual variables. In a re-analysis of Milroy's (1980) Belfast data using cluster analysis and PCA, backing of /a/ was not found to occupy a special position. Backing of /a/ is, however, an important sociolinguistic variable in Belfast, where it is used predominantly by working class men and is now being slowly adopted by women (Milroy & Milroy 1985b: 356-357). Thus, the change in progress identified by Milroy & Milroy on the basis of a careful item-based analysis escaped the multivariate analysis which as an exploratory technique grants equal weight to all variables. Methodological diversity is therefore important for sociolinguists working in the quantitative paradigm.

---

24 The question of whether or not to weight variables has been controversial with regard to biological classifications. While Sokal (1986: 426) emphasised the exploratory nature of clustering methods and rejected a priori weighting, Inger (1958) argued in his critique of numerical taxonomy that assigning greater weight to more important variables is central to the process of taxonomy.

25 While the 'old' uni- and bivariate as well as the 'newer' multivariate methods are readily available to sociolinguists, attempts should also be made to integrate new developments into the research methodology, such as, for example, fuzzy set theory and possibility theory which allow for a non-probabilistic modelling of non-random behaviour (see Smithson 1988, McNeill & Freiberger 1993: 70, Zimmerman 1991: 109-127, on possibility theory in particular see Klir 1989, for further comments on fuzzy set theory see § 8.1.1; see also Girard & Larmouth 1993, who suggested fuzzy set theory for the description of dialect transition areas).
PART III

DESCRIPTION AND INTERPRETATION OF THE CORPUS (1880-1922)
PATTERNS OF MORPHO-SYNTACTIC VARIATION

Boswell: Sir Alexander Dick tells me he remembers having a thousand people in a year to dine at his house ...
Johnson: That, Sir, is about three a day.
Boswell: How your statement lessens the idea.
Johnson: That, Sir, is the good of counting. It brings everything to a certainty which floated in the mind indefinitely.

J. Boswell The Life of Samuel Johnson (1791)

6

6.1 The Verbal System

6.1.1 Loss of Person and Number Distinctions in the Present Tense Paradigm

In (standard) Dutch the present tense paradigm for main verbs has three distinct forms: verb stem, stem plus inflectional -t and stem plus inflectional -e(n) (pronounced as a schwa; see Geerts et al. 1984: 431-434, 441-446, Booij & Van Santen 1995: 70-75).

(1)     infinitive: werk-e(n)
     1st sg.     werk-∅
     2nd/3rd sg.      werk-t
     1st/3rd pl.   werk-e(n)
     2nd pl.     werk-e(n), werk-t

In written Dutch, inflectional -t is not added to verbs with a stem ending in -t (zitte(n) → hij zit, wete(n) → hij weet). Verbs with stems ending in -d, do add the -t in the written language; this, however, does not affect their pronunciation (rijde(n) → hij rijdt). Inflectional -t is dropped in inversion in the 2nd person singular (je leest een boek → lees je een boek?). The 2nd person plural variant ending in inflectional -t is today marginal, and the form in -e(n) is commonly used (Donaldson 1981: 113, fn.).

In addition there exists a small set of five high-frequency, monosyllabic main verbs

---

1 In this chapter inflectional morphology is defined in accordance with Anderson's (1982: 587) 'rule of thumb' as 'what is relevant to syntax'. It is, however, acknowledged that to divide morphology into 'lexical' (derivational) and 'syntax-oriented' (inflectional) morphology, although conceptually useful, is problematic and that important counter-examples to the theory of a 'split-morphology' exist (see, for example, Van Marle 1996).

2 Final -n can be heard in front of vowels and in what is called 'reading pronunciation' (leesuitspraak), as well as in some regional varieties (Geerts et al. 1984: 424, Debrabandere 1997). However, since apocope of -n in unstressed syllables is typical in spoken Dutch, both -en and -e can be seen as scribal representations of the infinitive/plural marking of the verb.
(gaan, staan, slaan, doen and zien) and their derivatives (such as verstaan or aanzien), whose infinitive and plural forms do not end in -e(n) but in -n. Otherwise their conjugation is regular.

(2) infinitive: gaan
   1st sg.   ga-o
   2nd/3rd sg.  gaan-t
   1st/2nd/3rd pl.  gaan-n

The present tense conjugation of the auxiliary zijn has suppletive forms rather than inflections in the singular, while the paradigm of the auxiliary hebben is inflectional (albeit irregular in the 3rd person singular). Note that zijn has a second infinitive form (wezen) which replaces zijn in double infinitive constructions (such as in het zou erg leuk kunnen wezen or ik ben wezen kijken, Donaldson 1981: 140, Geerts et al. 1984: 444, 578). The forms bent/hebt are variants of the 2nd person plural, but less common than zijn/hebben (Donaldson 1981: 137, 139).

(3) infinitive: zijn (wezen)/ hebben
   1st sg.   ben/ heb
   2nd sg.   bent/ hebt
   3rd sg.   is/ heeft
   1st/3rd pl.  zijn/ hebben
   2nd pl.   zijn/ hebben; bent/ hebt

In spoken, informal Dutch the forms je heb/ ben and hij heeft/ heb are often heard, and with the second person formal plural pronoun u both hebt and heeft, as well as bent and is, are acceptable in continental Dutch.

The Dutch modal auxiliaries (with the exception of willen) belong to a historical class called 'Preterite Present' verbs, which are inflected like the preterite of strong verbs in the present tense (De Vooys 1953: 127-129, LeRoux & LeRoux 1973: 173).

(4) infinitive: mogen
   1st/2nd/3rd sg.  mag
   1st/3rd pl.  mogen
   2nd pl.   mogen, mag

For zullen and kunnen a variant ending in -t exists for the 2nd person singular (je zult, kunt), and varies with the invariant singular form (je zal, kan). For the 2nd person plural two forms exist, with the monosyllabic variant being less common than the variant in -e(n). Willen, although not a 'Preterite Present' verb, follows a similar inflectional paradigm, i.e. 3rd person singular without inflectional -t, and two forms for the 2nd person singular (je wil/ wilt) (Talen, Kollewijn & Buitenrust Hettema 1908: 114f., Geerts et al. 1984: 442-446).

---

3 Talen, Kollewijn & Buitenrust Hettema (1908: 112f.), De Vooys (1953: 132-134), Geerts et al. (1984: 442, 444); with respect to the acceptability of these forms in 19th century South African Dutch see Elffers (1893: 111f.).
Compared with Continental Dutch the verbal system of Afrikaans is characterised by a drastic reduction of the number of forms, resulting in the loss of person and number distinctions. The forms of the Afrikaans present tense paradigm are based on either the invariant verbal stem (main verbs), a present tense singular form (hebben/ zijn and the modal auxiliaries) or the plural/ infinitive form (verbs of the set 'gaan'). The Afrikaans form of the auxiliary 'to have' is based on the 3rd person singular variant het/heet (< heeft), which was common in 17th-century Dutch and still occurs in dialectal Dutch (Kloek 1956). Some relics of the old inflectional system exist in contemporary Afrikaans, but they are rare and regularised forms are generally common in the spoken language (Grayson 1962: 78, see also Scholtz 1963: 13, Ponelis 1993: 386/ 391ff.).

\[(5)\]
\[
\begin{align*}
\text{infinitive: werk} & \quad \text{infinitive: weer/ hé/het}^4 \\
\text{present, sg./pl.: werk} & \quad \text{present, sg./pl.: ts/heet} \\
\text{infinitive: gaan} & \quad \text{infinitive: mag} \\
\text{present, sg./pl.: gaan} & \quad \text{present, sg./pl.: mag}
\end{align*}
\]

Both internal and external factors have been advanced to explain the loss of number and person distinctions in Afrikaans. According to the position of the South African Philological School the Afrikaans present tense paradigm is the result of a gradual process of dialect levelling which was supported by interference from L2 varieties of Dutch. The process is believed to have been completed around 1750 (Raidt 1983: 127). 18th century CDV was characterised by high frequencies of concord disagreement, not all of which conformed to the emerging system of Afrikaans.

\[(6)\]
\[
\text{first person subject + 2nd/ 3rd person verb} \\
\text{ik noemt (Barbier, 1739)}
\]

\[(7)\]
\[
\text{sg. subject + plural verb/ infinitive} \\
\text{ik weeten (anon. 1720)}
\]

\[(8)\]
\[
\text{pl. subject + singular verb} \\
\text{de schapen is (H. Coetzee, 1748)}
\]

(examples from Ponelis 1993: 402-404, see also Griessel 1981: 43 for examples from the early 19th century)

Variation in the present tense paradigm has been reported for early Dutch (see Raidt 1983: 118) and is prominent in the Dutch dialects, which utilise a variety of inflectional systems,

\[4\] Also Van Loey (1959: 176), Scholtz (1963: 14); see Smits (1996: 130-131) for the use of het in Iowa Dutch.

\[5\] The infinitive form hé is used in constructions where the infinitive follows present or imperfect modal verbs (as in ek wil dit hé), while het is used as an auxiliary in compound tenses (as in hy moet die bok geskiet het; Donaldson 1993: 239-240).
illustrated for the singular paradigm in map 6.1.

Map 6.1 Dialectal present tense paradigms in the northern varieties of Dutch (singular only; Van den Berg 1949: 7).

Inflectional variation is no less pronounced in the plural. Stem plus inflectional -t, for example, has been reported for Drenthe, the eastern part of Overijssel and the east of Gelderland. Stem forms are today common in De Veluwe, and have also been reported for Zuthpen and Zelhem (Combrink 1978: 75-76, Smits 1996: 70-71, see also Van den Berg 1949, Hol 1955, Sayre 1972, Goeman 1976 [1983], 1993). Interacting with possible processes of dialect levelling in the early settlement was the tendency of West-Germanic languages towards loss of material in unstressed syllables, which supported the loss of verbal inflections (see Ferguson 1995 for a general discussion of loss of agreement as an example of 'drift' in Germanic, also Kastovsky 1994: 145-146).  

6 Phonetic attrition in posttonic syllables has commonly been attributed to the adoption of initial expiratory stress (Meillet 1970: 53-55, Kastovsky 1994: 145-146). Within the Germanic family, English has lost all verbal inflection except for the 3rd pers. singular. Swedish and Norwegian (bokmål) also do
The Afrikaans present tense paradigm has also been described as the result of contact-induced restructuring leading to the typical pidgin/creole feature of an unmarked verb (Valkhoff 1966: 25, Den Besten 1989: 235-236, Poneles 1993: 339, De Kleine 1997). Proponents of this position have emphasised not only the relative speed of the process (according to the datings of the South African Philological School less than 100 years), but also the fact that the Afrikaans verbal system is not really a straightforward example of inflectional loss; rather it is the result of different simplification processes. That is, the fact that some verbs continue Dutch stems, others Dutch infinitives or Dutch third person singular forms, is seen as suggesting reanalysis of morpheme boundaries by non-native speakers. The use of invariant verbs in basilectal varieties was furthermore supported by possible substrate influence since neither Khoe nor Pasar Malay or Creole Portuguese inflect verbs for person and/or number. That the process of verbal reduction was not limited to inflectional loss (i.e. material in unstressed syllables) but involved a general tendency towards deflection and regularisation, is acknowledged by Raidt who agrees that universals of second language acquisition ('Spracherwerb der Immigranten, Sklaven und eingeborenen Khoir'; 1983: 118) interacted with dialectal tendencies in a complex dialect levelling process. However, Raidt explicitly denies the possibility of a 'plötzliche Vereinfachung infolge einer Sprachenkonfrontation und Kommunikationskrise (Pidgin-Situation)' (ibid.: 126), and maintains that interference from L2 varieties only accelerated an essentially evolutive and gradual process of normal linguistic change (ibid., also 1991: 206-207).

That loss of person and number distinction need not be the catastrophic result of intensive language contact is suggested by the data for New Netherlands Dutch which existed in a prolonged but (initially) 'quite limited' contact situation with English, yet shows clear signs of morphological reduction (Van Marle & Smits 1993: 321, 323-324). However, comparative

not distinguish person and number in the present tense; however, they maintain the finite-infinite distinction. Loss of inflectional material is also common in spoken German whose standard (written) system is characterised by comprehensive inflectional paradigms. Thus, the full pronunciation of -en in expressions such as wir kommen is often lost in speech, where the verb is realised as /kom/ (see Lütke 1989: 135). However, adoption of initial stress and phonetic attrition are not universally related, since, for example, Icelandic, Finnish, Hungarian and Czech show initial stress without reduction of inflectional suffixes.

7 There seems to be some disagreement regarding the degree of morphological reduction in New Netherlands Dutch. Buccini (1992) summarised the situation as follows: 'While verbal inflection was reduced [...] the reduction was largely phonologically motivated: the principle remained until the end' (my emphasis; also Buccini 1996: 44). However, Van Marle & Smits (1993) argued that while Mohawk Dutch shows little morphological change, New Jersey Dutch shows clear signs of morphological reduction: Prince (1910) described the marking of person/number as variable, while Van Loon (1939) reported its complete disappearance (reported in Van Marle & Smits 1993). Van Ginneken (1913: 287-288), on the other hand, remarked that general loss of inflection was prominent only in the varieties spoken by exslaves, and Dillard (1976: 32) thus interpreted these varieties as creolised. New Netherlands Dutch is an
evidence from Germanic languages suggests that language contact and bilingualism generally support and possibly accelerate simplificatory changes (see § 2.2.3; for the possibility that contact contributed to loss of inflectional material in Germanic, see Kastovsky 1994: 150). Note also, that although verbal inflection codes meaning (i.e. person and number), it is essentially redundant in Dutch, which is not normally a pro-drop language.

The corpus (1880-1922) comprises over ten thousand finite present tense forms (N=11231). In order to focus on the emerging system of Afrikaans as distinct from Dutch, forms which are identical in both systems were excluded from the analysis as their systemic identity is ambiguous (i.e. they cannot be classified unambiguously as representing either system). The following categories were, therefore, excluded:

(i) 1st sg. of main verbs (Dt. *ik werk* - Afrik. *ek werk*), as well as 2nd and 3rd sg. of main verbs whose stem ends in -t/-d (i.e. which either do not inflect in the singular or whose inflection is not reflected in speech)
(ii) sg. of the modal auxiliaries (Dt. *ik zal, je zal/zult, hij zal* - Afrik. *ek, jy, hy zal*)
(iii) 3rd sg. for *hebben* / *zijn* (Dt. *hij heeft* (dialectal *het*); *is* - Afrik. *hy het; is*)
(iv) plurals and infinitives of the verbal set *'gaan'* (Dt. *wij gaan* - Afrik. *ons gaan*)

This left 5118 finite verb forms (1296 *hebben*, 871 *zijn*, 494 modal auxiliaries, 1905 main verbs and 552 verbs of the set *'gaan'*), and over half of these (54%) showed morphological reduction in accordance with the emerging system of Afrikaans and disagreed with the system of standard Dutch.

The degree to which the individuals in the sample used a verbal system conforming to the emerging norms of Afrikaans can be summarised by using a block histogram representing a grouped frequency distribution (figure 6.1). The input data are the relative percentages for each individual, i.e. the percentage of *'Afrikaans'* forms of their total number of finite verb forms (excluding the four categories discussed above). The height of the blocks in the diagram is proportional to the number of individuals in each class interval.

---

important point of reference, since (as Bucconi 1996 has shown) the social and regional origin of the settler population was very similar in the two settlements; however, the settlers in New Netherlands had only sporadic contact with non-European groups and the ratio of slaves to settlers was much smaller than in the Cape.
Figure 6.1 Histogram for 136 cases, summarising the degree to which the verbal system of individuals shows loss of person and number distinction (all verbs, based on percentages for individuals).

The distribution of cases (individuals) across data values is polarised as well as continuous. Two large clusters are situated on either side, that is, most individuals conform closely either to the system of standard Dutch (i.e. maintain the person/number distinction with few exceptions), or to the emerging system of Afrikaans (i.e. almost complete loss of the person/number distinction). Nevertheless, variable usage is pronounced with just under 50% of individuals in the middle field of the distribution.

Separating the data according to verb class (table 6.1) shows that (statistically significant) differences exist between verb classes.

<table>
<thead>
<tr>
<th></th>
<th>modal</th>
<th>hebben</th>
<th>main</th>
<th>zijn</th>
<th>'gaan'</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afrikaans forms</td>
<td>295</td>
<td>746</td>
<td>1065</td>
<td>402</td>
<td>228</td>
</tr>
<tr>
<td>Total</td>
<td>494</td>
<td>1296</td>
<td>1905</td>
<td>871</td>
<td>552</td>
</tr>
</tbody>
</table>

Afrikaans verb forms
as % of total | 60%  | 58%  | 56%  | 46%  | 41%    

Table 6.1 Use of Afrikaans forms according to verb class (C-1880-1922). $\chi^2 = 72.77769$, df.=9, $p < 0.001$.

The quantitative data given in table 6.1 translate into a smooth, yet rather flat S-curve, illustrating the spread of the change (loss of the person/number distinction) across environments (verb classes). The numerical differences between modals, the auxiliary hebben and main verbs,
however, are marginal when compared with the more considerable gap between this group and
*zijn* 'gaan'. Separating the singular and plural of main verbs did not reveal any (statistically
reliable) patterns (table 6.2). However, a significantly lower frequency of uninflected forms
occurred in the first person plural (48%, 179 of 371, \( \chi^2 = 10.64144, \text{df}=3, p=0.01 \)).

<table>
<thead>
<tr>
<th></th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verb + ( \sigma )</td>
<td>705</td>
<td>360</td>
</tr>
<tr>
<td>total</td>
<td>1231</td>
<td>674</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>verb+ ( \sigma ) cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>as % of total</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>57%</strong></td>
</tr>
</tbody>
</table>

| **53%**                  |

Table 6.2 Loss of verbal inflection for main verbs (C-1880-1922). \( \chi^2 = 2.691929, \text{df}=3, p=0.4 \).

In her analysis of inflectional change in Iowa Dutch, Smits (1996: 121) found that in the
early sample (collected in 1966) the use of uninflected forms was still marginal in the plural of
main verbs, which she interpreted as an attempt to maintain surface regularity.

The marginal occurrence of verb stems in the plural may be related to the fact that in that context
the generalization of monosyllabic and consonant-final verb stems affects surface regularity.
Prototypical plural verb forms, namely, should be polysyllabic and end in -en. The use of verb
stems in the 2nd and 3rd person singular, however, does not affect surface regularity, since
prototypical singular forms are monosyllabic and end in a consonant.

However, some 20 years later (1989) the use of uninflected verbs had spread into the
plural paradigm of Iowa Dutch (Smits 1996: 204-205). Similarly Franken (1953: 171), notes in
his analysis of the Duminy diary (see § 2.1) that stem-forms are rare in the plural but common
in the singular, which suggests that inflectional loss in (acrolectal) CDV also started in the
singular, then spread to the 2nd and 3rd person plural, and finally to the 1st person plural (which
as mentioned above showed a significantly lower frequency of uninflected forms). Use of
uninflected verbs in the plural possibly affected first the 2nd person plural which in (standard)
Dutch has a monosyllabic variant ending in -t. Monosyllabic 2nd person plurals were also
prominent in some older varieties of Dutch and are still typical in dialectal Dutch (Scholtz 1963:
19, Goeman 1993: 115). In the corpus all 2nd person plural forms were uninflected; however,
since only very few tokens were available, the evidence for assuming a special status of the 2nd
pers. pl. is rather meagre. The spread of monosyllabic stem forms to the 3rd person plural could
have been triggered by the fact that the Dutch third person plural pronoun *zij/ze* is homonymic
with the female 3rd person singular pronoun; however, this explanation is problematic since in
continental Dutch the two forms are always kept apart.

Variation within the paradigm of main verbs is only marginally due to relic forms derived
from Dutch plurals/infinitives or 2nd/3rd person singular forms which compete (even in modern Afrikaans) with stem forms. According to Poneiris (1993: 393) -e forms were common until three generations ago, but their use has been greatly reduced since. Variation between stem forms and stem plus -el-t affects verbs with f/v or d/t voicing alternation (skryf ~ skrywe, bid ~ bidde), /x/ syncope (Dt. klagen → kla'en → Afrik. kla → klae), and verbs with stem final resonant (reën ~ reënt, begin ~ beginn; Scholtz 1963: 12-14, Poneiris 1993: 389). The relative frequency of uninflected vs. (relic) inflected forms is furthermore lexically constrained, and is influenced by social and contextual variables such as region, age and speaker preference (Poneiris 1993: 391-393). The fluid and often idiosyncratic nature of the variation patterns of these verbs was already emphasised by Malherbe (1917: 55, see also Hoogenhout 1904: 13).

The data summarised in Table 6.3 seem to support the interpretation that variability within this group was partially lexically motivated. However, the data values in Table 6.4 show that the same was true for verbs with stem-final voiceless obstruents which in Afrikaans only occur as stem forms. This suggests that loss of verbal inflection in CDV might (in general) be amenable to a lexical diffusionist explanation.

<table>
<thead>
<tr>
<th></th>
<th>schrijven</th>
<th>leven</th>
<th>regenen</th>
<th>beginnen</th>
</tr>
</thead>
<tbody>
<tr>
<td>verb - ø</td>
<td>17</td>
<td>6</td>
<td>6</td>
<td>21</td>
</tr>
<tr>
<td>total</td>
<td>48</td>
<td>15</td>
<td>20</td>
<td>41</td>
</tr>
</tbody>
</table>

Stem forms as %
of total: 35% 40% 30% 51%

Table 6.3 Loss of verbal inflection for *schrijven*, *leven*, *regenen*, and *beginnen* (C-1880-1922). $\chi^2 = 14.98368$, df=7, $p<0.04$; expected frequencies were calculated on the basis of the average for main verbs.

<table>
<thead>
<tr>
<th></th>
<th>maken</th>
<th>spreeken</th>
</tr>
</thead>
<tbody>
<tr>
<td>verb - ø</td>
<td>32</td>
<td>6</td>
</tr>
<tr>
<td>total</td>
<td>45</td>
<td>22</td>
</tr>
</tbody>
</table>

Stem forms as %
of total: 71% 27%

Table 6.4 Loss of verbal inflection for *maken* and *spreken* (C-1880-1922). $\chi^2 = 11.0100$, df=3, $p<0.01$; expected frequencies were calculated on the basis of the average for main verbs.

The auxiliaries *hebben* and *zijn* show a pattern different from that of the main verbs when
separated according to number (table 6.5 and 6.6). That is, regularisation is more prominent in the plural than in the singular, but only for hebben is the distribution statistically significant. In the paradigms for both zijn and hebben a significantly smaller number of reduced forms is found in the 1st person pl., which parallels the patterns observed for main verbs (for zijn 30%, 42 out of 140, $\chi^2 = 16.64574$, df=3, p<0.001; for hebben 59%, 170 out of 289, $\chi^2 = 5.520053$, df=1, p=0.02).

<table>
<thead>
<tr>
<th></th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>is</td>
<td>155</td>
<td>247</td>
</tr>
<tr>
<td>total</td>
<td>347</td>
<td>524</td>
</tr>
<tr>
<td>use of is as % of total</td>
<td>45%</td>
<td>47%</td>
</tr>
</tbody>
</table>

Table 6.5 Generalisation of is (C-1880-1922). $\chi^2 = 0.4818981$, df.=3, p = 0.9.

<table>
<thead>
<tr>
<th></th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>het/ heeft</td>
<td>444</td>
<td>302</td>
</tr>
<tr>
<td>total</td>
<td>833</td>
<td>463</td>
</tr>
<tr>
<td>use of het/ heeft as % of total</td>
<td>47%</td>
<td>65%</td>
</tr>
</tbody>
</table>

Table 6.6 Generalisation of het/ heeft (C-1880-1922). $\chi^2 = 17.20940$, df.=3, p<0.01. Note than het/heeft were treated as allomorphs for the analysis, re. the use of het see § 7.2.

Although examples of the new system abound in the corpus, non-Afrikaans concord disagreement (as described for 18th century CDV) still occurs.

(9) \textit{first person subject + 2nd/ 3rd person verb}  
\begin{itemize}  
\item ik vraagt (Wilhelmina January, 30/3/1897)  
\item ik denkt (Cornelia Stanford, 6/8/1905)  
\end{itemize}

(10) \textit{sg. subject + plural verb/ infinitive}  
\begin{itemize}  
\item dat de Doctor hunne voeten snyden en reg draai (B.J. Brümmer, 21/6/1900)  
\item ik moeten (Mozes Mlimkulu, 1/12/1896)  
\end{itemize}

(11) \textit{pl. subject + singular verb}  
\begin{itemize}  
\item motor cars komt (Wyanda Hoogenhout 18/2/1907)  
\item ik en mijn echtgenoote ben (MJ. Heroldt, circa 1892)  
\end{itemize}

However, in terms of quantity, such examples of 'misuse' of inflection (for the term see Smits 1996: 43) are negligible (265 instances, i.e. 2% of all finite verb forms), and less than half of the sampled individuals show examples of non-Afrikaans concord disagreement (58 out of 152 cases). Most of these examples involve variation in the verbal set 'gaan' (see below), or relics forms such as skryve; archaic use of -t in the 1st person singular is also relatively common. There is, however, a small group of four individuals (Beatrice de Waal, Wilhelmina January, Mozes
Matodlana, Jacobus November) who show percentages of between 15% - 30% of misuse of inflection, including the frequent use of -e(n) with singular verbs which is otherwise extremely rare. The linguistic behaviour of this small group is best described as hypercorrection, i.e. an attempt to counteract the emerging (and possibly stigmatised) new system by using verb forms which were clearly marked as being inflected, yet their imperfect knowledge of the Dutch standard prevented their use of the correct forms of the old system.⁸

Regarding the emergence of the new system for the verbal set gaan, staan, doen, staan and zien, non-Afrikaans non-concord forms are rather more common than in the other verbal categories. The form ending in -t (gaat) was frequently generalised to the 1st singular and the 1st/2nd and 3rd person plural (35 out of 192, 18%), while the stem form (ga) occurred only occasionally in the plural or the 2nd/3rd person singular (19 out of 547, 4%; see Changuion 1844: 121, Pannevis 1880 [1971]: 78, Mansvelt 1884 [1971]: 158).⁹ Variation between -n and -t is still known in contemporary Afrikaans for this set of verbs (Ponelis 1993: 391, Donaldson 1993: 218).

(12) maar ik ziet geen kans (B.J.Brümmer 26/4/1902) ons gaat almal naar StilBaai (J.F.W.Grooskopf 8/12/1907)

(13) [de] paarden die men hier zie is groot maar te lang van rug (L.Botha 12/8/1909) aanstaande week ga Danie Mev. Retief en kinderen halen (Annie van Huyssteen 23/7/1907)

(14) liefste Boykie, jy zien ik is by Annie (Johanna Brümmer, 10/5/1895) als mogelyk komt voor hy gaan (F.M.C.Breuris 22/1/1906)

If one accepts the traditional dating for the loss of person and number distinctions in the present tense paradigm (i.e. around 1750, Raidt 1983: 127, 1991: 217), then the variability in the corpus (1880-1922) is surprising, and it is questionable whether it can be explained merely as a result of dialect contact under diglossic conditions (i.e. contact between CDV and standard Dutch) as argued, for example, by Scholtz and Raidt (see § 2.1). A similar line of thinking was

⁸ It has been pointed out by Janda & Auger (1992) that the term 'hypercorrection' is currently used to described two different processes: (i) quantitative hypercorrection, and (ii) qualitative hypercorrection. Quantitative hypercorrection refers to the observation that members of a lower status groups show a greater frequency of prestige/standard variants in formal situations than do members of higher status groups (the classic illustration is the crossover pattern in Labov's /h/-graph for NYC 1991[1972]: 114). Qualitative hypercorrection, on the other hand, refers to (frequently isolated) cases where speakers (or writers) become self-conscious about their linguistic performance resulting in the misapplication of linguistic rules which speakers consider prestigious (see also Pargman 1998 on strategies of hypercorrection and their role in language change).

⁹ Generalisation of stem forms is prominent only in the letters of Louis Botha (7 out of 12, 58%) and Petronella van Huyssteen (3 out of 22, 14%).
already suggested by N.M. Hoogenhout (1904: 13) in his grammar of Afrikaans.

In Gebeten usw. versucht man ausnahmslos seine Sprache, so gut oder so schlecht es geht, der niederländischen Schriftsprache anzupassen. Daher eine außerordentlich bunte Formenmischung. (Hoogenhout 1904: 13)

In other words, the church and an increasingly elaborated school system brought the members of the Netherlandic speech community into (relatively) regular contact with the norms of continental Dutch leading to linguistic insecurity, random variability and ad hoc dialect mixture (via borrowings from the standard language) in situations which were characterised as 'formal' (including written registers). However, this interpretation is problematic on two grounds:

(i) as discussed above the variability was not random but quantitatively constrained by verb class membership (modals > hebben > main verbs > zijn > 'gaat'). To a lesser extent differences between singular and plural are noticeable, and the 2nd person plural still lagged significantly behind for main verbs as well as zijn and hebben. There is also some indication that different lexical verbs participated to differing degrees in the process.\(^\text{10}\)

(ii) 'misuse' of inflection or hypercorrection is marginal in the corpus as a whole, and absent from the linguistic production of the majority of individuals.

In a recent paper on the absence of hypercorrection in East Anglian present tense verb forms, Trudgill (1996) argued that the absence of hypercorrect behaviour indicates that the observed variation patterns are probably not the result of the unstabilising and random influence of the standard language but constitute 'a permanent and continuing feature' of the language or dialect (1996: 110). Although hypercorrect behaviour is not absent from the corpus, it involves mostly relic forms such as skrywe and gaat, and is otherwise limited to a small group of individuals. In general the corpus is characterised by patterns of variable reduction of verbal marking, which suggests that rather than a completed change by 1750, the pace of change in the present tense system was relatively slow and continued well into the late 19th and possibly even early 20th centuries. Whether this gradual pattern was limited to acrolectal/mesolectal CDV (see § 2.3) is a separate question, and outside the scope of this study.

6.1.2 Re-visiting Conradie (1979)

The apcope of [t] - An example of morphophonemic variation: The loss of verbal agreement in main verbs for the 2nd and 3rd person singular interacted with the phonological

---

\(^\text{10}\) Note that lexical diffusionist explanations have been linked to speaker-centred accounts of linguistic change as discussed in chapter 1; thus Chen 1977: 217 stated: 'The word represents the level of consciousness par excellence of speakers' use of language. To exclude the lexical dimension is to ignore the role of speakers' active participation in linguistic change.'
process of t-apocope, common not only in CDV but also in 17th century Dutch and today's Dutch dialects. T-apocope has been reported for upper class varieties of 16th and 17th century Dutch, and its prominence in CDV suggests that Dutch prestige forms played a role in the development of Afrikaans (Snyman 1979, Stoops 1980, Ponelis 1993: 154-155). Word-final [t] deletion is still common in many dialectal varieties of Dutch (especially Holland and Utrecht; De Vries 1974, Goeman 1983 [1976]).

Within the verbal system t-apocope not only affected the finite verb (2nd/3rd person singular present tense) but also the (regular and irregular) weak past participle. In Dutch the weak past participle has a discontinuous inflection: ge-[stem]-t/d (-t after voiceless consonants, -d after voiced consonants, realised as [t] due to final devoicing, e.g. jagen → ge-jaagged, koken → ge-kook-t).11 Afrikaans has maintained the prefix ge- as a past participle marker, but lost the ending -t/d (kook → gekook). In both Dutch and Afrikaans the prefix ge- does not precede atomic verbal prefixes (thus ontmooten → onmoet, beloven → beloof(d)).12 However, regularised forms such as gebegin or gegegen have been reported for Griqua Afrikaans and other non-standard varieties (see Rademeyer 1938: 62-63, Loubser 1954: 102, Tiffin 1984: 62, Webb 1993: 163). Only very few of such forms were found in the corpus.

(15) die wordt hem geontnomen (Anton April 6/7/1913)

T-apocope in CDV was a variable rule in the Labovian sense as certain phonological and grammatical factors furthered or inhibited the application of the rule, i.e. [+voice] of the preceding segment as well as word class (finite verb, PP or non-verbal item; Scholtz 1963: 15-17, Conradie 1979: 117-123, Lubbe 1983). In his variationist study of the diachrony of the past participle in Afrikaans, Conradie analysed the morphophonemic conditioning of t-apocope quantitatively by dividing his data (17 texts or text complexes ranging from 1739-1851) according to the following categories:

(i) 3rd sg. present tense forms with stems ending in [-voice]
(ii) past participles with stems ending in [-voice]
(iii) non-verbal items ending in [-voice]
(iv) 3rd sg. present tense forms with stems ending in [+voice]
(v) past participles with stems ending in [+voice] (Conradie 1979:92; 1981, 1994)

---

11 In Dutch three types of past participles (and thus verbs) are distinguished: weak and regular (koken → ge-kook-t), weak and irregular (with stem mutation, kopen → ge-koch-t) and strong (with ablaut, schrijven → ge-schrev-en).

12 Afrikaans also allows omission of the ge-prefix in double infinitive constructions where the main verb is followed by an infinitive: Dit het ophou reën (< opgehou reën; Donaldson 1993: 225).
The overall pattern of Conradie's analysis shows clearly that t-apocope is more common after voiceless consonants than after voiced consonants; it is also more frequent in the 3rd person sg. present tense than in the past participle and non-verbal items, the latter never showing t-apocope after [+voice]. Regarding the development of the morphophonemic variation pattern, Conradie distinguished on the basis of his analysis three stages.\textsuperscript{13}

**Stage A**
The phonological factor supports t-apocope after [-voice] in all word classes, while the grammatical factor supports t-apocope in the 3rd person sg. present tense only.

**Stage B**
Largely a continuation of stage A but the grammatical factor now also supports t-apocope for past participles. This leads to a small increase of t-apocope in the past participle especially after [+voice].

**Stage C**
The phonological factor no longer plays a role for t-apocope in the verbal system (present tense and past participle). T-apocope is, however, still more common in the 3rd person present tense form than in the past participle.

![Graphs showing the development of t-apocope](https://via.placeholder.com/150)

**Figure 6.2** The development of t-apocope according to phonological environment and grammatical constraint (adapted from Conrad 1979: 127-129). MVPT = minus voice present tense; MVPP = minus voice past participle; MVNV = minus voice non-verbal; PVPT = plus voice present tense; PVPP = plus voice past participle.

\textsuperscript{13} Notice that these three stages are typologically rather than chronologically successive. The texts which conform to stage A were written between 1797 and 1838, stage B 1829-1851, stage C 1830-1854 (see Conrad 1979: 130). Ponelis (1993: 155) is pessimistic with respect to the possibility of understanding the actual chronological development: 'The early Afrikaans data does not permit chronological ordering of the contexts of [t] apocope in order to determine the spread of the process.'
Regarding the further development of the variation pattern, Conradie predicts that in a later stage t-apocope in the past participle will catch up with the present tense paradigm (1979: 129), thus leading to the generalisation of t-apocope for inflectional [t]. Finally loss of [t] after [-voice] will be generalised for non-verbal items. Non-verbal items in which a voiced consonant precedes [t] remained unaffected in (standard) Afrikaans (but see Ponelis 1993: 155 for historical and dialectal examples of t-apocope after resonants in non-verbal items).

Replicating Conradie's approach, frequency of t-apocope was analysed for 12 individuals who had sufficient tokens (according to the criteria outlined in § 4.2) for all five environments. The overall pattern of the data (using mean scores) lies somewhere between Conradie's stage B and stage C, i.e. the grammatical factor is prominent, supporting t-apocope for present tense as well as increasingly for the past participle, yet still interacts with the phonological factor of [±voice].

![Figure 6.3 Apocope of [t] according to phonological and grammatical constraints (based on percentages, mean scores for 12 cases).](image)

A closer analysis of the individual graphs shows the existence of a subset of individuals whose behaviour reflects a pattern similar to the one predicted by Conradie, i.e. the frequency difference between 3rd person sg. present tense and past participles has almost disappeared. However, interestingly, the same is not true for the phonological constraint which (although losing its force) still supports frequency differences depending on the nature of the preceding consonant. This pattern in most clearly visible for Johanna Brümmer (but also underlies the linguistic behaviour of G.B. Goosen, J.W.F. Grosskopf, B. Smit and Wynanda Hoogenhout [2]).
This pattern should not be interpreted as following stage C, which was already characterised by the loss of the phonological constraint; rather it suggests the possibility that after Conradie's stage B the diachronic development bifurcated into one pattern C[1] (Conradie's C) in which the phonological constraint of [±] voice disappeared but the grammatical constraint (present tense vs. past participle) was maintained. A second pattern C[2] was characterised by the continuation of the phonological constraint while the grammatical constraint affected present tense and past participle identically. The possible development of t-apocope can be visualised as follows:

![Diagram](image)

**Figure 6.5 Possible lines of development of t-apocope.**

A very advanced pattern, which could be termed stage D (if such a label is felt necessary), is evident in F.S. Malan's performance, where t-apocope is nearly categorical.
Figure 6.6 Apocope of [t] according to phonological and grammatical constraints: F.S. Malan (based on percentages).

An almost perfect example of Conradie’s stage C[1] (loss of the phonological constraint) is found in the linguistic behaviour of Petronella van Huyssteen (figure 6.7), i.e. maintenance of the grammatical constraint but loss of the phonological constraint. The graphs for Aletta Schabort and Carolina Leipoldt (figure 6.8 and 6.9) can be interpreted as approaching stage C[1], however, Leipoldt still maintains final /t/ categorically in the past participle after voiced consonants (a relic from phase A), while the phonological factor still plays a considerable role in the linguistic behaviour of Schabort.

Figure 6.7 Apocope of [t] according to phonological and grammatical constraints: P. van Huyssteen (based on percentages).
Figure 6.8 Apocope of [t] according to phonological and grammatical constraints: C. Leipoldt (based on percentages).

Figure 6.9 Apocope of [t] according to phonological and grammatical constraints: A. Schabort (based on percentages).

B.J. Brümmel's language use (figure 6.10) shows increasing loss of the phonological constraint in the present tense, but maintenance in the past participle. Non-verbal items show an unusually high frequency of t-apocope. The most conservative pattern is found in the letters by Louis Botha (figure 6.11): t-apocope is considerable only in the present tense, where it is strongly affected by the phonological constraint.
Figure 6.10 Apocope of [t] according to phonological and grammatical constraints: B.J. Brimmer (based on percentages).

Figure 6.11 Apocope of [t] according to phonological and grammatical constraints: Louis Botha (based on percentages).

The graphs show that variation patterns identified by Conradie for the late 18th/early 19th century were still productive around 1900, suggesting, in opposition to the conventional view which sees the process as being completed before 1800 (Raidt 1983: 80-82, Ponelis 1993: 155), continuing variation or a process of slow linguistic change.
Apocope of -e(n): Unlike t-apocope, loss of word-final -e(n), which affected within the verbal system the plural forms of the present tense paradigm, the infinitive and the past participle of strong verbs, was not the result of a phonological process. According to Scholtz (1963: 17-37) apocope of -e(n) gradually spread through the grammatical categories leading to inflectional loss in the plural of the main verb by 1750, the infinitive around 1800, and finally the past participle (in predicative function) by the mid 19th century (Raidt 1983: 124-127). During the last quarter of the 19th century a formal distinction arose between predicative and attributive past participles, with the latter maintaining the inflected form (Conradie 1979, 1982).

(16) breek → hy het die blompot gebreek
die gebroke blompot

The maintenance of the (original) inflected form in secondary functions is an example of Kuryłowicz's fourth law of analogy: 'When a form undergoes differentiation as a consequence of a morphological change, the new form corresponds to its primary function and the original form is reserved for the secondary function' (quoted in Anderson 1992: 370).

According to Conradie's (1979: 173) analysis of 18th and early 19th century data, frequency of apocope of -e(n) was roughly similar in the infinitive and past participle, while apocope in the present tense plural was far ahead of both (average for all texts: present tense 77%, infinitive 17% and past participle 19%). The quantitatively marginal ordering of past participle > infinitive, noticeable in Conradie's data, was not found to exist in the corpus. Instead one observes the following (statistically significant) cline: verb > infinitive > PP, which supports Scholtz's original hypothesis of a grammatical constraint.

<table>
<thead>
<tr>
<th></th>
<th>plural verbs (present tense)</th>
<th>infinitive</th>
<th>past participle (predicative)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-e(n) loss</td>
<td>635</td>
<td>1639</td>
<td>358</td>
</tr>
<tr>
<td>total</td>
<td>1168</td>
<td>4129</td>
<td>1104</td>
</tr>
<tr>
<td>-e(n) loss as % of total</td>
<td>56%</td>
<td>40%</td>
<td>32%</td>
</tr>
</tbody>
</table>

Table 6.7 Apocope of -e(n) for plural verbs (present tense), infinitives and past participles (C-1880-1922). Following Conradie (1979: 171) forms of hebben/ zijn were excluded. χ²=144.9338, df= 5, p< 0.001.

Regarding -e(n) apocope in the infinitive, an important difference was found to exist between infinitives occurring in infinitive clauses (te and om...te infinitives; on the infinitive clause in Dutch and Afrikaans, see §7.3.2), and infinitives following finite auxiliary verbs (modals as well as copulas such as gaan and laten). Infinitives in infinitive clauses showed a stronger tendency to maintain the inflectional ending than those following a finite modal or copula. This pattern reflects the principle formulated by Scherre & Naro (1992: 1-2) according
to which 'marking leads to more marking', i.e. overt marking of the infinitive clause by the preverbal particle *te* (and possibly the complementizer *om*) can be said to trigger inflectional marking of the verb.

<table>
<thead>
<tr>
<th>infinitive clause</th>
<th>following a modal or linking verb</th>
</tr>
</thead>
<tbody>
<tr>
<td>-e(n) loss</td>
<td>484</td>
</tr>
<tr>
<td>total</td>
<td>1431</td>
</tr>
<tr>
<td>-e(n) loss as % of total</td>
<td>34%</td>
</tr>
<tr>
<td></td>
<td>1155</td>
</tr>
<tr>
<td></td>
<td>2698</td>
</tr>
<tr>
<td></td>
<td>43%</td>
</tr>
</tbody>
</table>

Table 6.8 Apocope of -e(n) for infinitives (C-1880-1922). $\chi^2 = 31.7896$, df = 3, $p < 0.001$.

The past participle of strong verbs in Dutch is not only characterised by the inflectional ending -e(n), but also by ablaut which was lost in Afrikaans. 14

(17) Dutch: spreken $\leftrightarrow$ gespreken
Afrikaans: spreek $\leftrightarrow$ gesprek

An important question regarding the development of the regularised system of Afrikaans is thus whether the inflectional ending and ablaut were lost at the same time, or whether one of them was lost prior to the other, that is, whether there was an intermediate stage (Scholtz 1963: 35).

\[ \text{gespreken} \rightarrow \text{gesproken} \rightarrow \text{gespreek} \rightarrow \text{gesprook} \]

Figure 6.12 Possible lines of development for the regularisation of the past participle of strong verbs.

The historical evidence indicates that the path of development was different for those past participles with a back vowel as ablaut vowel in the past participle (group I; $<i>$, $<e>$ $\leftrightarrow$ $<o>$ and $<\delta>$, $<ui>$, $<ie>$ $\leftrightarrow$ $<\delta>$), and those with a front vowel (group II; $<ij>$, $<i>$ $\rightarrow$ $<\delta>$; Conradie 1981: 281). In the corpus 259 examples of group I past participles, and 231 examples of group II past participles were found and tabulated according to the following categories:

14 Den Besten (1989: 35) noted that unlike the loss of inflectional material, loss of ablaut should not be attributed to dialect levelling, since *[a]*blaut leads a tenacious life in Dutch and its dialects'. He attributes the loss of ablaut to gradual influence from a stable Cape Dutch pidgin/creole which had lost the ablaut entirely (see § 2.2.2).
A both ablaut and -e(n) maintained  
B ablaut lost, but -e(n) maintained  
C ablaut maintained, but -e(n) lost  
D both ablaut and -e(n) lost  

\[
\begin{array}{cccc}
\text{no. of tokens} & A & B & C & D \\
\text{% of total} & 213 & 2 & 1 & 43 \\
\end{array}
\]

Table 6.9 Regularisation of the past participle, group I (\(<i>, <e> \rightarrow <o>\) and \(<i>, <u>, <ie> \rightarrow <o>\) (C-1880-1922). N= 259.

\[
\begin{array}{cccc}
\text{no. of tokens} & A & B & C & D \\
\text{% of total} & 126 & 28 & 4 & 73 \\
\end{array}
\]

Table 6.10 Regularisation of the past participle, group II (\(<ij>, <i> \rightarrow <e>\) (C-1880-1922). N= 231.

The percentages given in table 6.9 and 6.10 support Conradi's (1981) suggestion that regularisation followed different paths in group I and II, and that verbs with a front ablaut vowel in the past participle were regularised earlier than those with a back vowel. Only group II shows a notable percentage of past participles which have lost the ablaut but maintained the inflectional ending (which is also true for Conradi's analysis). This can be interpreted as an indication that only within group II regularisation contained an intermediate step. Verbs belonging to group I possibly maintained the ablaut vowel when group II lost it, but (perhaps by analogy with the forms from group II) lost at a later time both ablaut and inflectional ending simultaneously.15

At the same time the frequency of reduction was again to some extent lexically constrained. The distribution for the verb *geschreven* (N= 104), for example, shows only 15% for the loss of both ablaut and -e(n), while loss of ablaut but maintenance of the inflectional ending was observed in 26% of cases. For *krijgen* (N=78), on the other hand, 51% (N= 40) of forms were found to conform to the regularised paradigm of Afrikaans (loss of ablaut and loss of -e(n)), and there is only one instance of category B (loss of ablaut, maintenance of -e(n)). In group I *nemen* (N=71) and especially *beginnen* (N=9) show a very high percentage of regularised forms.

15 It should be noted that Pheiffer's (1980) study of the interlectal Dutch of the French Huguenots suggests a different path of development for group I. Of the 107 cases in which ablaut was maintained but -e(n) was lost only 25 cases belong to group II, which suggests that in the interlectal varieties of the Huguenots verbs belonging to group I lost the ablaut after the loss of -e(n) (see also Conradi 1981: 281).
Table 6.11 Regularisation in past participle for *nemen*, *beginnen*, *schrijven* and *krijgen* (C-1880-1922).

Conradie’s analysis is a convenient point of reference, since his investigation of the history of the past participle followed a quantitative and variationist approach which also forms the basis of this study. A comparison of his results with the variation patterns around 1900 (i.e. roughly a century later) suggests a situation of continuing variation, rather than completed language change. On the other hand, while the overall pattern remained stable, the relative frequencies increased, indicating that change (i.e. inflectional loss) continued, although possibly at a slower pace than traditionally assumed by historians of Afrikaans.

6.2 Tense

6.2.1 Past Tense Variation

Continental Dutch has a grammaticalised distinction between preterite and perfect which was lost in Afrikaans. In Dutch the preterite of main verbs is expressed by verbal inflection, the perfect by a periphrastic construction (inflected auxiliary verb + past participle). The preterite of regular verbs is marked by the ending *-te* (after voiceless segments) or *-de* (after voiced segments) in the singular, and *-te(n)*/*-de(n)* in the plural.\(^\text{16}\) The preterite of strong irregular verbs is marked by a change in stem vowel (ablaut), and the ending *-en* in the plural. Weak, irregular verbs inflect according to the paradigm for regular verbs (*-te(n)*/*-de(n)*), but also show stem mutation (Geerts et al. 1984: 430, Booij & Van Santen 1995: 71-72).

\(^{16}\) Since apocope of final *-n* is common in continental Dutch (see fn.2) plural preterite verbs are generally unmarked for plural in spoken Dutch.
(18) **Preterite Tense (Preterite).**

I. Regular verbs
   
   | Dutch | Dutch
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>koken &gt;</td>
<td>ik kook-te, wij kook-ten</td>
</tr>
<tr>
<td>horen &gt;</td>
<td>ik hoor-de, wij hoor-dcn</td>
</tr>
</tbody>
</table>

II. Irregular verbs
   
   a. weak irregular
      
      | Dutch | Dutch |
      |-------|-------|
      | kopen > | ik kocht, wij kocht-en |
   
   b. strong irregular
      
      | Dutch | Dutch |
      |-------|-------|
      | zingen > | ik zong, wij zong-en |

**Present Perfect**

<table>
<thead>
<tr>
<th>Dutch</th>
<th>Dutch</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. koken &gt;</td>
<td>ik heb ge-kook-t/ ge-hoor-d</td>
</tr>
<tr>
<td>II.a. kopen &gt;</td>
<td>ik heb ge-kocht-t</td>
</tr>
<tr>
<td>II.b. zingen &gt;</td>
<td>ik heb ge-zong-en</td>
</tr>
</tbody>
</table>

**Past Perfect (Pluperfect)**

<table>
<thead>
<tr>
<th>Dutch</th>
<th>Dutch</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. koken &gt;</td>
<td>ik had ge-kook-t/ ge-hoor-d</td>
</tr>
<tr>
<td>II.a. kopen &gt;</td>
<td>ik had ge-kocht-t</td>
</tr>
<tr>
<td>II.b. zingen &gt;</td>
<td>ik had ge-zong-en</td>
</tr>
</tbody>
</table>

The semantic distinction between preterite and perfect is difficult to pin down, and is subject to considerable variation in the West Germanic languages. A rough guide to the semantics of tense and aspect can be found in Reichenbach's (1947: 287-298, see also Seuren 1996: 84-86) analysis of tense which distinguishes three relevant types of time.

- **S** time of speaking
- **R** reference time, i.e. the time established within the discourse/text
- **E** event time, i.e. the time of the event which is being described

The preterite is commonly used to describe past situations or actions which took place prior to the present (time of speaking), while reference time and event time coincide: $S \succ R = E$ ($\succ$ stands for 'posterior to'). Regarding the internal contour of the action (aspect) the preterite is said to indicate duration, habituality and/or continuity.

(19) De kinderen speelden beneden.
    Iedere dag ging hij naar de bibliotheek.

The semantic function of the perfect can be described in terms of 'completion' and 'anteriority', i.e. the perfect refers (a) to a completed action ('perfective aspect'), which (b) took place some time prior to the reference time established within the discourse. The present perfect is used in cases where the reference time lies in the present (and thus coincides with S), while the action/situation described is located prior to the present: $S = R \succ E$. 
(20) Ik lees niet graag, maar dat boek heb ik gelezen.

The past perfect (or pluperfect) is used in cases where the reference time itself is located in the past, and the reported event occurred prior to the reference time (S>R>E).

(21) Hij sliep nadat hij de krant gelezen had.

In contemporary Dutch (as well as in German) the present perfect is used increasingly for the narration of past events, and has usurped many of the functions originally covered by the preterite (for more detailed discussion see Bosker 1961, Comrie 1985: 82). In both Dutch and German, for example, the present perfect can be used for the description of a past event which has only perfective aspect, but which does not fulfill the criterion of anteriority. In English past tense would be required in this situation (see Vogel 1997: 91-92).

(22) Ik ben naar binnengegaan en heb de inbreker genegenheid.

Afrikaans has lost the preterite forms as well as the past perfect, and retained the present perfect as the only marked past tense form (ek het gehoor, ek het gesing). There are few relics of preterite forms in Afrikaans. These are mostly auxiliaries (kan/kon, mag/mog, moet/moes, sal/sou, wil/wou, wees/was), rarely used are the preterite forms of weet (wis), dink (dag, dog) and hé/het (had) (Ponelis 1993: 429ff. Donaldson 1993: 222-224).\(^\text{17}\) According to the chronology established by the South African Philological School the past perfect was lost in Afrikaans around 1775, the preterite followed in the first half of the 19th century (Scholtz 1963: 37-4, Raidt 1983: 121-123, 127). Scholtz (1972: 31) has argued that the loss of the preterite was a slow and gradual process, which was completed only fairly late, and possibly affected different sociolinguistic varieties at different times.\(^\text{18}\) Note also that the question of whether the preterite should form part of the emerging standard of Afrikaans was still controversial in the early 20th century (see Steyn 1989: 20-21).

Regarding the explanation of the development, both evolutive language change and creolisation have been suggested. Lass has argued that the loss of the grammaticalised past/perfect distinction in Afrikaans 'is actually part of a recurrent or cyclical process in Germanic

---

\(^{17}\) The modal auxiliaries have been regularised in KVA. Kotzé gives the following examples for 'Malay Afrikaans': Hy kon sien > Hy kan gestien het and ek wou al die kinders stuur > ek wil al die kinders gestuur het (1989: 260). The preterite variants of dink originate from standard and dialectal Dutch respectively (standard denken- ik docht, dialectal denken - ik docht; Donaldson 1993: 22, fn.8).

\(^{18}\) See for example Scholtz's comment with respect to the use of preterites by Bezuidenhout (1851): 'Het ons hier te make met verskillende lae van Afrikaans, of kon Bezuidenhout hom net nie so goed van die Ndl. skryfradisie losmaak nie?' [my emphasis] (1963: 46).
as well as other branches of IE' (1987b: 326, see also Lass 1997a: 255-256). Thus, the past(aorist)/perfect distinction of Proto-IE was lost in Proto-Germanic, but reintroduced in Germanic times when a new grammaticalised perfect (AUX + PP) was created. In Yiddish and Afrikaans (also reported for Upper German from as early as the 15th century) perfect and preterite have now merged again: the preterite forms have been lost and the perfect has taken over the functions of the old preterite (for Yiddish see Weinreich 1992: 319-320, for German see Von Polenz 1991: 199, Erben 1980: 97-98, see also Visser 1992: 2-12). Ponelis (1993: 439, also De Kleine 1997), on the other hand, although acknowledging possible language internal causes, has described the loss of the preterite in Afrikaans primarily as a result of contact-induced restructuring, i.e. 'another aspect of verbal deflection resulting from verbal invariance under creolisation'.

In the corpus variation between preterites and present perfects shows a clear preference for the perfect (main verbs only, 1766 instances of present perfects out of a total of 2344 past tense forms excluding pluperfects, 75%), i.e. a past tense ratio of roughly 2:1 in favour of the perfect. Early Modern Dutch and contemporary Dutch dialects, on the other hand, show a preference for the preterite (63% in Early Modern Dutch and 75% in contemporary Leiden speech, see Ponelis 1993: 430-431). According to Ponelis' (1993: 448-449) quantitative analysis of 18th century CDV data a preference for the perfect was discernible quite early at the Cape (63% for the 1709-1754 period, and 74% for the Kaapse Taalargief). In the corpus, an important difference exists between main verbs and the auxiliaries hebben/zijn, the latter of which occur only rarely as periphrastic present perfects. The fact that the auxiliaries maintained two past tenses (i.e. a preterite form and a perfect form) was noted in the 19th century by Du Toit (1976 [1897]: 25), however, in modern Afrikaans only was is commonly used (and varies with is gewees) while had (rather than het gehad) is rare even in writing (Donaldson 1993: 236-240).

<table>
<thead>
<tr>
<th>main verbs</th>
<th>hebben/ zijn</th>
</tr>
</thead>
<tbody>
<tr>
<td>no. of perfect forms</td>
<td>1766</td>
</tr>
<tr>
<td>total forms</td>
<td>2344</td>
</tr>
<tr>
<td>perfect forms as % of total</td>
<td>75%</td>
</tr>
</tbody>
</table>

Table 6.12 Present perfect-preterite ratio (C-1880-1922). $\chi^2=1300.525$, df= 3, $p<0.001$.

The stem-and-leaf plot illustrating the distribution of cases (individuals) across data values (main verbs only), does not show a broad variation continuum with two clearly identifiable poles (as for the verbal system), but rather an exponential pattern with over two-thirds of individuals showing scores of over 60%.
The distribution of individuals indicates that the change in the past tense system (unlike, for example, the loss of person and number distinctions) had penetrated the acrolectal usage of the entire speech community, with only few individuals maintaining a past tense ratio in favour of the preterite. Note that hypercorrect preterite forms are non-existent in the data; the only incorrect preterite found being *ik Jong* (Dt. *ik ging*, J.P.Z. Marais 9/11/1918), formed in analogy with other /-g/ verbs (such as *zingen - ik zong*). Examples of past perfect still occur but are rare. However, one individual (Cornelius Coopman) stands out in that he seems to have generalised past perfect forms with *had* as a marker of past tense. A similar tendency has been reported for Griqua Afrikaans (see Rademeyer 1938: 63-64, Roberge 1998: 32; see also Elffers 1903: 18), where the auxiliaries *het* and *had* are used interchangeably. Another (albeit sporadic) characteristic of Griqua Afrikaans, namely, the omission of the tense auxiliaries *hebben/zijn* (see Roberge 1998: 27), was absent from the corpus.

### 6.2.2 Tense and Aspect Adverbials

In Afrikaans tense adverbials are often used to specify the temporal and aspectual dimensions of an utterance (Donaldson 1993: 228ff.). The tense adverb *al 'already',* for example, can be used to indicate anteriority, and occurs in sentences where the past perfect would occur
in Dutch.\footnote{Upper German which also lost the past perfect, has developed a 'double perfect' as a substitute: 'Zufällig habe ich an diesem Tag eine Forelle gestohlen gehabt, und der Fischer ist zornig zu uns gelaufen und hat geschrien.' (Erben 1980: 98). A similar construction exists in Afrikaans (\textit{Hy het dit toe al gedoen gehad}) but is rarely used (Donaldson 1993: 233).}

(23) Voordat dit begin reën het, het sy al reëls haar reënjas aangetrek. (Donaldson 1993: 232)

Adverbs are also used to specify perfective aspect. In Afrikaans the Dutch form of the past participle of \textit{doen} (\textit{gedaan}, rather than the Afrikaans variant \textit{gedoen}) can be used as an indicator of completive aspect (often in combination with \textit{al}; see Roberge 1994b: 77–78, 1998: 33, for a similar construction in AAVE and its multi-syncretic history, see Schneider 1993). The oldest examples for (\textit{al}) \textit{gedaan} as a perfective marker (resultive perfect) are recorded in the 18th century Khoe and slave data, thus suggesting the possibility of a basilectal origin of this construction.

(24) Die gift \textit{al gedaan} dood, wie kan hy meer wat schaden (Khoi-speaker 18th century) (quoted in Roberge 1994b: 78)

\textit{die Clercq} heeft geseg, \textit{ijt mijn Camraat gedaan} vast maken (1720) (quoted in den Besten 1989: 225)

Completeive aspect can also be expressed by the adverb (\textit{al}) \textit{klaar} (Dutch 'ready, finished', Afrikaans 'ready, already'), which according to the interpretation of Den Besten (1989) continues the pattern of (\textit{al}) \textit{gedaan}. The adverbs \textit{al} and \textit{alreeds}, both meaning 'already', can also be used in this function.


\textit{Toe(n)} which can function both as an adverbial ('then') and a complementizer ('when') is increasingly used with present tense verbs in Afrikaans (also attested for 18th century CDV, see Scholtz 1972: 30). In Dutch \textit{toen} ('then', 'when') requires the use of past tense forms (Donaldson 1981: 105; 122, Geerts \textit{et al.} 1984: 462, 652).

(26) \textit{Toe} ('when') hy daar aankom, \textit{toe} ('then') groet die ou omie hom. (Donaldson 1993: 228)

The use of the unmarked verb (present tense) to describe past events/situations is widespread in Afrikaans, and occurs also within the corpus where it often functions as a discourse marker. The following extract comes from a letter by Rudolph Baalie in which he describes a rather unpleasant journey in an over-booked train. Using the preterite tense he describes in the first part of the narrative how he found a seat for his wife and younger children;
he then ventured to find some space on the train for himself and his older children (bold font: preterite, caps: present perfect, italics: present tense).

(27) Tot hierheen waren wij tezamen in een carriage, toen moesten wy ons verdelen, de vrouw ging met de kleinere kinders in een carriage waar reeds 2 vrouwen en 1 kind in waren en ik ging met de groteren in eene carriage waar 1 police met 2 kaffer-convicts in waren. Die 2 vrouwen, die van Johannesburg kwamen, en lekker sliepen waren natuurlijk zeer boos om zoo gestoord te worden, en wonden ook niet opstaan. Zoo moest de arme Marie met de kinders [...] de nacht doorbrengen in eene zeer onvriendelijke gezelschap.
Met mij ging het nog erger, de police joeg mij weder uit, ik liet de kinders zitten en ga naar die station-master, en zeg hem, hij HEEFT GEZEID: ik moet hierin en de police jaagt mij uit. Hij kwam zamen en vraagt de police waaron moet die man uit? Hij zegt: Ik was de eerste hierin en ik ben met convicts hier, ik wil hier geen andere menschen in hebben. Goed zegt die station master; Ik moet alle mensen die tickets hebben plaats geven, er is geen andere plaats meer dan hier. Ga in! zegt hij voor mij en als hij U weder uitjaagt, then you wire back, and I shall teach him. (Rudolf Baalie 23/2/1894)

In this extract the switch to the present tense functions as a discourse marker, separating the anecdote (which includes long stretches of direct speech) from the rather straightforward description of the train journey. The switch to the predominant use of present tense in example (27) is very much in line with what we know from tense variation in oral discourse, where present tense was found to be the stylistic norm when narrating comic stories and anecdotes (on tense variation and discourse marking see Wolfson 1979, Leith 1995; see also the comments made by Franckenh 1900). A similar pattern of variation also exists in Aletta Schabors't memoir where passages involving reported speech and passages involving general statements (when, for example, describing the dirt or the heat in the camp) typically favour a switch to the present tense.

Proponents of a possible creole origin of Afrikaans have paid much attention to the widespread use of the (unmarked) historical present as well as to the existence of non-verbal aspect markers (Den Besten 1989: 236, 238, De Kleine 1997: 302-304). The frequent use of the historical present has been explicitly linked to the systems of creole languages.

Whereas [Dutch] employs the historical present as a stylistic device, the use of a present tense form of the verb in a past tense context in [Afrikaans] looks more like a pattern known in some of the creole languages: tense, once established in discourse, may be left unmarked. (Bruyn & Veenstra 1993: 40)

Extensive usage of the historical present is, however, also possible in English. As mentioned above Leith (1995) has argued that the historical present can function as the norm in narratives, and Wolfson's (1979) analysis has shown that tense variation is often part of the
organisation of the narrative, and is thus best approached from the perspective of discourse analysis (see example 27). The use of the historical present in Afrikaans, although possibly more widespread than in other Germanic languages, is thus by no means 'un-Germanic' (as suggested by De Kleine 1997: 302; note also that tense marking can be obligatory in Afrikaans, see Conradie 1998: 37). The same applies to the use of adverbs to specify aspektual and temporal information within the discourse, which is too common in non-Creole languages to be taken as an indicator of creole origin (see, for example, Crystal 1966 on the use of tense adverbials + unmarked verb in English, and Cook 1992 on adverbial aspect marking in German). Elicited information from Afrikaans speakers suggests furthermore, that rather than a clear and stable system of tense and aspect formatives (as known for creole languages), the use of tense/aspect adverbs in Afrikaans is highly context-dependent and subject to much variation. Holm (1988/89 II: 347), although in general favourable towards those emphasising the creole character of Afrikaans, summarised the situation as follows:

[The Afrikaans verbal system remains quite Germanic despite its considerable morphological simplification. It is definitely not the verbal system of the Atlantic creoles, with preverbal tense and aspect markers.

In the corpus there are two examples of constructions in which gedaan has been lexicalised as a (predicative) adjective with the meaning 'exhausted, finished, used up' which according to Roberge (1995: 83) formed the basis for its use as a perfective marker.

(28) sommige zoorten pruimen is al reeds gedaan en de andere zoorten zyn nu ryp (BJBrümmer, 6/01/1901)
   ik het naderhand Bob in my kamer laat slaap [...], want Bob was te gedaan (Johanna Brümmer 152/2/1895)

While several instances of both al (perfective and anterior) and toe(n) (past) occur in the corpus, only few examples of klaar + verb indicating 'already' or 'finished' (perfective aspect) were found. There are two examples in the data in which klaar has been lexicalised as a predicative adjective meaning 'exhausted': de ossen is almal klaar gery ('the oxen are exhausted from the ride', J.G. Du Preez 11/8/1894), and zo velen oude leeraren, die klaar gewerk is ('so many teachers who are exhausted/ finished from their work', M.A.M. Oosthuizen 3/101/1906). This lexicalisation along the same lines as reported for gedaan can be seen as corroborative evidence for Roberge's (1995: 84) argument that perfective klaar 'continues the pattern (al)

Note, however, that the situation is different in some non-standard varieties of Afrikaans. Griqua Afrikaans, for example, has a system of relatively stable pre-verbal markers, such as the durative marker lê and the future marker loop (Rademeyer 1938: 78-79, Roberge 1994c: 58ff.).
gedaan in the Cape Dutch pidgin' (also Den Besten 1989: 238).

(29) (al) klaar 'perfective'
    Ik heef voor een week terug klaar geskeer (Eddie Smith, circa 1900)
    Ek geloof u Taalbond examens vragen zijn al klaar uitgeskreven (Bettie Smit 20/5/1917)

(30) al 'anterior' and 'perfective'
    ek het hulle ook al geseh dat hulle geen Afrikaans met oupa moet praat (Hester
    Hoogenhout, 10/11/1916)
    U het zeker al gehoor dat ik nou skool hou. (Leon Malan, 31/10/1917)

(31) al reeds 'anterior' and 'perfective'
    Als te zamen heb ik al reeds £4 ontvangen van de Heer Smidt. (M. Mlimkulu 1/12/1896)

(32) toe (n) (when, then) 'past'
    toen ik het geld zien huil ik hart (M. Hillebrand 13/10/1903)
    Gister aand toen ik hier aankom was prof Viljoen net uit (F. S. Malan 7/6/1896)
    we konden niet denken wat zij wou hebben + toe ze zij ik wil water (Jacomina van
    Huyssteen 14/4/1908)

Only 62 cases of toe(n) + present tense were found in the data (out of a total of 318
instances of toe(n), i.e. 20%). Poncelis' (1993: 432-34) quantitative analysis of toe(n) (both as
complementizer and adverbial) in an extract from the Lewende Bybel (published 1983) indicates
a much greater preference for the use of the historic present (in 25 cases (total: 38) toe
was followed by the present tense, i.e. 66%). It is thus possible that the generalisation of toe(n) +
present tense is a development of the 20th century, since today toe(n) is 'most usually' used with
present tense (Donaldson 1993: 228).

In sum, although aspects of the system of tense adverbials probably continue earlier
baselectural structures, the use of tense adverbials is in general rather variable and highly context-
dependent, and is thus in line with the use of tense adverbs in other West-Germanic languages
such as German and English. Widespread use of toe + present tense, which has been cited as
evidence for creolisation, is possibly a 20th century development; however more (quantitative
and qualitative) work is necessary to settle this question.

6.3 Grammatical Gender

6.3.1 Loss of Gender Agreement

Continental Dutch has a two-gender system which distinguishes common gender (de-
words, comprising both historically masculine and feminine words; roughly two-thirds of Dutch
nouns) and neuter gender (het-words) in the singular (for details see Geerts et al. 1984: 41-52).
The old IE three-gender system is maintained only in the Flemish dialects (Dekeyser 1980: 106-
108). The Dutch system of de- and het-nouns is difficult to acquire for L2 speakers as it largely
lacks overt morphological or semantic motivation, but it is of central importance for the
morphosyntactic structure, as the gender of the head noun determines the form of definite articles and demonstratives, relative pronouns, anaphoric and deictic pronouns, as well as the inflection of attributive adjectives.

Grammatical gender was lost completely in Afrikaans, where the invariant definite article die is used with all nouns. Gender irregularities occurred in L1 as well as L2 texts from the late 17th century (Scholtz 1963: 125ff., 1972: 39). Raitd has dated the loss of gender to 1750 (Raitd 1983: 149, 1991: 208; see also Scholtz 1963: 125), while Ponelis has argued that gender was still relatively intact in the second half of the 18th century (Ponelis 1993: 172). As agreement is the way in which gender is realised syntactically, loss of gender can be defined as the partial or complete loss of agreement (Corbett 1991: 105, 315). The quantitative analysis of loss of gender in the corpus takes lack of agreement in two different types of determiners as a diagnostic feature for loss of grammatical gender: the definite article (de/het) and the demonstrative article (deze, die/ dit, dat). To combine definite articles and demonstratives was necessary, since in Afrikaans die can function as both article and demonstrative, and it is often impossible to determine from the written text whether a particular example of die represents a definite article or a demonstrative. Two linguistic strategies can be seen as indicating the process of gender breakdown: (i) the use of common determiners with neuter singular nouns, and (ii) the use of determiners marked for neuter with common nouns (singular).

1166 historically neuter singular nouns were tabulated, 637 of which were preceded by common determiners (55%). The opposite strategy, i.e. neuter article or demonstrative with common gender noun, was negligible (less than 1%). As in the case of verbal agreement the distribution of cases has two clearly identifiable poles (i.e. loss of agreement < 10% and almost complete loss of agreement), and between the poles the distribution is continuous (45% of individuals are found in the middle field of the distribution).

---

21 Note, however, that the gender system of Early Modern Dutch was not entirely stable and some high frequency nouns fluctuated between common and neuter gender (Ponelis 1993: 132).

22 Case marking, which was still widespread in 19th century written Dutch, has disappeared in both modern Dutch and Afrikaans. There are, however, relics of case inflection in some Afrikaans expressions, such as: dag der dae (genitive plural) or in der haast (dative singular; for more examples see Grayson 1962: 24ff.). In the corpus case marking is rare.
Figure 6.15 Histogram for 60 cases, summarising the loss of neuter gender across cases (based on percentages for individuals).

Neuter nouns used with a common determiner include high-frequency nouns such as *geld, jaar, veld, and huis*. High frequency is usually interpreted as a conservative force which can inhibit grammatical reduction (Hooper 1976, Bybee & Brewer 1981, Stemberger & MacWhinney 1986, Smits 1993), and loss of gender in these lexical items shows that the reduction process was deeply entrenched in 19th century CDV.

<table>
<thead>
<tr>
<th>common determiner de/die</th>
<th>huis</th>
<th>geld</th>
<th>jaar</th>
<th>veld</th>
<th>dorp</th>
</tr>
</thead>
<tbody>
<tr>
<td>total de/die % of total</td>
<td>57%</td>
<td>30%</td>
<td>20%</td>
<td>18%</td>
<td>21%</td>
</tr>
</tbody>
</table>

Table 6.13 Loss of neuter gender in five high frequency nouns (C-1880-1922). Note that this distribution is not statistically significant, $\chi^2 = 10.06880$, df=9, p=0.3.

In Dutch (as well as in other Germanic languages) there exist a number of suffixes which function as morphological gender markers. These include *-heid* and *-ij* for common gender (*de schoonheid, de boekerij*) and the diminutive *-je* (with the variants *-kie, -pije, -tie* and *-etje*) for neuter gender (*het kastje; see Donaldson 1981: 27-32, Geerts et al. 1984: 42-46*). The diminutive ending had clearly lost its function as a signal for neuter in CDV at the turn of the century, and over two-thirds of the diminutives in the corpus occur with a common article or determiner (64
cases out of 92, i.e. 70%). Interestingly this process appears not to be determined by the semantic gender of the noun. Carolina Leipoldt, for example, uses *de Janneke* [+human, +male], *de kannetje* [-human], *de tuitje* [-human], but *het meisje* [+human, +female]. *Meisje* is an interesting case, as it constitutes an example of what Corbett (1991: 225) has termed 'hybrid nouns', i.e. nouns in which semantic meaning and grammatical gender conflict with each other, leading to variation in gender agreement. Such variation is usually not random, but follows what Corbett (1991: 226) has called the 'agreement hierarchy':

**The Agreement Hierarchy**

attributive < (predicative) < relative pronoun < personal pronoun  
(Predicative does not require agreement in West Germanic)

The further right an element is placed in the hierarchy, the more likely it is that agreement will be motivated by semantic factors. Thus, as it has been described for the German neuter noun *Mädchen*: agreement is common in attributive agreement targets (*das/dieses nette Mädchen*) and relative pronouns (*das nette Mädchen, das...*), while (anaphoric) personal pronouns often show semantic and not grammatical agreement (*das nette Mädchen, das .... Es/Sie ist nett*). The situation in Dutch is similar, and neuter nouns referring to male or female persons often take pronouns reflecting the semantic gender of the antecedent.

(33)  
*Waar is dat meisje ineens gebleven?*  
*Ik weet het niet, ik heb haar niet gezien.*

An important characteristic which distinguishes the (anaphoric) personal pronoun from other agreement targets is the fact that it can be widely separated from its antecedent, often across sentence boundaries. Its reference function in language is, furthermore, twofold: it refers to the grammatical antecedent as well as to the referent denoted by the antecedental noun. In cases where the natural gender of the referent clashes with the grammatical gender of the referential noun, anaphoric pronouns are at the forefront of the conflict between the semantic and grammatical identity of hybrid nouns, and are therefore of special importance for the description of changes in gender systems (Corbett 1991: 247). Examples of hybrid nouns denoting male or female referents and semantically motivated (anaphoric) personal pronouns are common in the corpus and occur frequently in texts which otherwise conform closely to the norms of continental Dutch.

---

23 The frequency difference between diminutives and other neuter nouns regarding loss of gender agreement in the determiner is statistically significant: $\chi^2 = 8.174241$, df=2, $p=0.04$. That is, diminutives occur more frequently with a common determiner than neuter nouns in general.
(34) *het meisje*
't Meisje van Algie laat me zooveel aan haar denken. Die begint ook nou met loopen (Elsie van Huyssteen 2/3/1907)

(35) *het dochtertje*
Hij drukte het dochtertje ettelijke malen aan zijn boezem, kuste haar + ging (Algie van Huyssteen, 27/7/1909)

(36) *het zoontje*
Myn zuster Koenraad het ook een zoontje [,] het is die eerste (Cornelia Stanford 6/8/1905)

Loss of gender agreement with hybrid nouns is also found in determiners and relativisers (for the latter see § 7.1.3.).

(37) als ik in de kerk trouwd waar de meisje behoord (P.J. Eksteen, 15/3/1899)
een kleine meisje van omtrent 15 jaren die gaarne na U komen wil (C. Leipoldt, 4/9/1880)

Another important hybrid noun in the corpus is the neuter noun *het kind*, whose referents are [+ human] and can be both male and female. Just as with diminutives denoting male or female referents, pronominal agreement for the neuter noun kind is often semantically based.

(38) en hoewel de lieve kind nu niet zo weli is, maar in bed met pijn in de linker zijde (Anna De Vries, 11/6/1906)
en pas mijn kind mooi op[,] wees een moeder over hem (J. Henderson, c. 1900)
dat ik noch geen kind heb die voor gezondheid van huis moet gaan (M. Basson 22/6/1904)

Gradual loss of the gender distinction is a common fate of extraterritorial varieties of Dutch: it has been reported for New Netherland Dutch and East Indian Dutch, and the gender distinction is said to be 'weak' in Surinamese Dutch and Antillean Dutch (Van Marle 1995). Reduction of the gender system is, furthermore, typical for the diachronic development of the West Germanic languages, where it is one of the processes leading to the gradual 'loss of older West Germanic structures or contrasts' (Lass 1987b: 323). Continental Dutch itself, which has lost the old tripartite gender system (which is still maintained in Yiddish, German and the Flemish dialects of Dutch), has moved in the direction of reduction, but only English and Afrikaans have lost gender entirely (see also Ferguson 1995: 185). As with most changes in Afrikaans we are most likely dealing with a combination of different factors, most importantly Germanic 'drift' and language contact, both of which support (as mentioned in §2.2.3) simplificatory changes.
6.3.2 Attributive Adjective Inflection

The attributive adjective inflection in Dutch is governed by the categories of gender, definiteness and number. Singular neuter head nouns favour -o (unless preceded by an definite determiner), common gender and plural nouns the inflectional ending -e (pronounced as a schwa). In continental Dutch it is thus the inflected adjective which constitutes the 'general case' (Van Marle 1995: 284; also De Rooij 1980: 112), while the undeclined form is limited to one type of environment: [+neuter, -plural, -definite] (Booij & Van Santen 1995: 78).

(39) a. de jong-e man
dei jong-e meisje
b. (een) jong-e man
(een) jong-o meisje
c. de jong-e mannen
de jong-e meisjes

Adjectives denoting materials and those ending in -a, -o, -e, -i, -y and -en (including past participles used as adjectives) are not inflected (de nylon regenjas, het lila handdoek, een geschreven brief). Adjectives ending in -er are usually inflected unless they are derived from a geographical place name (Edammer kaas; Geerts et al. 1984: 321-322). The system of adjectival inflection in Dutch is, furthermore, subject to social, stylistic and geographical variation. Thus, the type "definite article + adjective + singular (neuter or common) head noun" typically shows the inflected form of the adjective but uninflected forms are possible, especially in constructions which form a semantic unit: het doctoraal examen, het bijvoeglijk naamwoord, de waarnemend burgermeester (Donaldson 1981: 75, Geerts et al. 1984: 328, Odijk 1992: 198-199). Neuter nouns preceded by a possessive pronoun also favour the uninflected form of the adjective in many varieties of spoken Dutch (haar zie kind, mijn nieuw boek; De Rooij 1980: 110, 113-119, Raitd 1983: 147, Smits 1996: 67). In indefinite common gender NPs where the head noun refers to a person the adjective can be used uninflected when used figuratively; a strategy 'exploited by Dutch speakers and writers for expression of a wide variety of subtle semantic nuances' (Shetter 1994: 60, Geerts et al. 1984: 329-330).

(40) een grote man a big man
een groot man a great man

With the gradual erosion of grammatical gender from 1700, a central trigger for the adjective inflection (i.e. gender of head noun) was lost, and the 18th century sources show

---

24 Note, however, that there appear to be restrictions on what lexical nouns can occur in this construction (*vent, *kerel, *jongen, see Odijk 1992: 198).
largely random variation (for examples see Scholtz 1963: 129ff.). However, instead of simply regularising the surface exponents of an oblique inflectional paradigm (always -ø or always -e) the system was restructured in Afrikaans in such a way that the inflection of the attributive adjective is determined no longer by morphosyntactic criteria but by the form of the adjective itself (lexicalisation).\(^{25}\) In Afrikaans polysyllabic and morphologically complex adjectives (with the exception of those ending in -er or -el) take the suffix -e categorically. Monosyllabic adjectives are uninflected, unless their (original Dutch) attributive form is strongly different from the simplex as a result of Afrikaans phonological processes such as apocope of /v/, lentition of /d/ or voicing alteration (sag > 'n sarge bed, goed > 'n goeie boek, blind [blint] > 'n blinde [blinde] kind; Raidt 1983: 142-147, Donaldson 1993: 163-169). Thus, Afrikaans maintained not only the statistically dominant inflected adjective (from the perspective of continental Dutch), but also the special case of the uninflected adjective.

As in Dutch where obligatory adjective inflection can be dropped to achieve a semantic distinction, non-inflecting adjectives are occasionally inflected in Afrikaans for semantic reasons.

(41) \begin{align*}
\text{die arm man} & \quad \text{i.e. the poor (penniless) man} \\
\text{die arme man} & \quad \text{i.e. the poor man, the man to be pitied (Ponelis 1993: 365)}
\end{align*}

Inflection can also be re-introduced to increase the degree of formality and is maintained in fixed expressions (as in die Verre Ooste). Attributive adjective inflection is still subject to variation in contemporary Afrikaans, and even native speakers often show insecurity about which forms are accepted as 'correct' (Carstens 1991: 81). In flux is, for example the inflection of monosyllabic simplexes ending in consonant + s. Some are categorically uninflected (spits, vars, pers, etc.), some categorically inflected (snaakse, slinkse, as well as those which are derived from a noun, as in die Kaapse klimaat), while others show variable inflection (fluks(e), preuts(e), rats(e); Ponelis 1993: 366). Regarding the group of monosyllabic adjectives ending in -g, there is disagreement among grammarians. While Donaldson classifies monosyllabic adjectives ending in nasal + g as inflecting, both Ponelis (1993: 366) and Raidt (1968: 107-108, 1983: 144) identify such adjectives (for example, streng or bang) as uninflecting.

Uninflected forms of monosyllabic adjectives are documented from the beginning of the 18th century, most frequently in the lexical items klein, groot and oud. Adjectives ending in -er

---

\(^{25}\) Regularisation was attempted by S.J. du Toit (see § 3.2) in his grammar of Afrikaans, when stating that Afrikaans categorically demands -e in attributive adjectives, while predicates occur without the suffix (1980 [1876]: 14, also Elffers 1903: 15). However, he frequently violated this rule himself (writing, for example, een groot letter and een rou boer). Inflectional generalisation has been reported for KVA and Western Cape Afrikaans (Kotzé 1989: 258; Ponelis 1993: 373, see also Hoogenhout 1904: 7).
also occur uninflected by the 18th century (Ponelis 1993: 367f.). Raidt (1968: 172-173, 1983: 152-153) has suggested that the maintenance of the special case of uninflected attributive adjectives in Afrikaans is partially due to patterns of adjectival inflection in upper-class varieties of 17th century Dutch (on the influence of upper-class varieties on CDV see also the discussion of t-apocope in § 6.1.2.). These varieties showed use of the uninflected adjective in additional environments: with indefinite determiner + masculine personal name in the nominative, and in cases where the adjective ends in -er. More recently, however, Van Marle has argued against assuming a direct link between the 17th century input and the current Afrikaans adjective inflection.

In my view it is very unlikely [...] that there is a direct link between the prominent position of the undecorated adjective in these varieties of overseas Dutch [Afrikaans and Old New York Dutch, the latter of which has generalised the uninflected adjective- A.D.] and the fact that in seventeenth century Dutch the undecorated adjective was more popular than in the present-day language. In the first place, there can be no doubt that in seventeenth-century Dutch, too, it was the declined adjective which represented the general case [...]. Secondly, the popularity of the undecorated adjective in seventeenth-century Dutch was primarily a characteristic of the cultivated language, and it seems very doubtful whether this latter variety of seventeenth century Dutch has been very influential in relation to the rise and development of Afrikaans and Old New York Dutch. (1995: 291)

Van Marle has attributed the generalisation of uninflected adjectives in Old New York Dutch to language contact with English, but does not advance an explanation for the system of Afrikaans. Lass (1990) has described the Afrikaans adjective inflection as an example of 'exaptation', i.e. co-optation of surface structures originally developed for other purposes to new functions, and characterised the development of a type of adjectival declension based on morphological (complexes vs. simplexes) and morphophonemic (monosyllabic adjectives with variant stem allomorphy and adjectives ending in -er/ -el) criteria as 'a conceptually novel inflection-type, with no real Germanic precedents' (Lass 1990: 91, also Van Marle 1995: 92: 'As a matter of fact, I am not aware of any other language with a system of adjectival declension which is even reminiscent of Afrikaans'). However, although the Afrikaans adjectival declension is a paradigmatic example of what Ferguson (1995: 190) has called an 'elegance innovation', that is an innovation 'that introduces a neat, pleasingly designed pattern that 'catches on' chiefly by virtue of its appeal to the human system-perceiving and system-constructing capacities', it is not entirely without Germanic antecedents, in this case a marginal (but widespread) tendency within Germanic to allow for a group of categorically undeclining adjectives (usually based on morphological or phonological criteria). Thus, as discussed above continental Dutch shows minor phonological/morphological (and even semantic) conditioning, by excluding certain types of adjectives (those denoting materials, those ending in vowel or -e(n), polysyllabic adjectives)
from the declension paradigm. In Swedish, adjectives ending in -s, -e, and -a are not inflected, Norwegian (bokmål) does not inflect adjectives ending in stressed vowel, unstressed -e or -s, and German does not inflect adjectives ending in vowel (das rosa Bändchen). In Frisian certain monosyllabic adjectives, such as äld, jong, lys and wij remain uninflected in cases where noun and adjective 'enter into a fixed collocation' (Hoekstra & Tiersma 1994: 513, as in de jong faam 'the young woman'). However, while the general principle (i.e. to exclude certain types of adjectives from the inflectional paradigm) is not unknown in other Germanic languages, Afrikaans is the only one of these languages which has transformed this marginal feature into a complex paradigm which is truly 'unique in Germanic' (Lass 1990: 95). The new declension paradigm is also remarkable in that its very structure does not parallel the general tendency towards reduction and regularisation so evident in other aspects of the morphosyntactic system. In other words, while gender and the verbal system went down the road of reduction and simplification, the declension of the adjective is an example of a complex re-structuring process. The attributive adjectives in the corpus (N=1766) were grouped into four categories:

(i) monosyllabic adjectives (categorically non-inflecting in Afrikaans) (N=566)
(ii) monosyllabic adjectives (categorically inflecting in Afrikaans) (N=382)
(iii) polysyllabic adjectives (categorically inflecting in Afrikaans) (N=712)
(iv) polysyllabic adjectives ending in -er (categorically non-inflecting in Afrikaans) (N=106).

To analyse the extent to which the new group of categorically non-inflecting monosyllabic adjectives was present in the corpus, only those environments where the adjective would have occurred as inflected in continental (standard) Dutch were tabulated. Of a total of 479 cases where Dutch requires inflected forms of the adjective, 203 did not have an inflectional ending (i.e. 42%). This overall result does, however, conceal once again important differences between different lexemes. Five high-frequency adjectives (arm, klein, groot, mooi and ou) were analysed separately and the results are summarised in table 6.13.


27 There were only four examples of polysyllabic adjectives ending in -el which were excluded from the analysis. Adjectives which are categorically uninflecting in Dutch (see above) were also excluded.
<table>
<thead>
<tr>
<th></th>
<th>arm</th>
<th>groot</th>
<th>klein</th>
<th>ou</th>
<th>mooi</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i)</td>
<td>6</td>
<td>36</td>
<td>22</td>
<td>70</td>
<td>26</td>
</tr>
<tr>
<td>(ii)</td>
<td>58</td>
<td>96</td>
<td>47</td>
<td>93</td>
<td>30</td>
</tr>
<tr>
<td>(iii)</td>
<td>10%</td>
<td>38%</td>
<td>47%</td>
<td>75%</td>
<td>87%</td>
</tr>
</tbody>
</table>

Table 6.14 Frequency patterns for five monosyllabic adjectives (C-1880-1922). (i) number of cases without an inflectional ending, (ii) number of loci where Dutch would have required an inflectional ending, (iii) percentage of cases without an inflectional ending, i.e. (i) as percentage of (ii) (χ²=82.52215, df=9, p<0.001).

The extremely low percentage of uninflecting forms of the adjective *arm* is probably due to the fact that almost nowhere in the data is *arm* used in the literal sense of 'penniless' but always in the transferred sense of 'to be pitied'.

(42) die arme mans (A. Schabort, 1902)
die arme dier (M. Marren, 13/6/1915)
on arme volk (P.G. Kuhn, 16/5/1900)

*Ou*, on the other hand, which shows a high percentage of uninflected forms, occurs uninflected even in the writings of those whose texts conform otherwise closely to the norms of standard Dutch, and is generally common in affective and informal contexts.

(43) ou broer (J.H.H. De Waal, 8/1/1902)
ou Katie (Rijkie Louw, 4/8/1909)
de ou boeken (F.S. Malan 1/6/1893)

Since, as mentioned above, uninflected adjectives are often found after possessive pronouns in dialectal Dutch, the effect of possessive determiners on loss on inflection was quantified in the data. 41 examples of monosyllabic adjectives (categorically uninflecting in Afrikaans, but requiring inflection in Dutch) preceded by a possessive pronoun were analysed, and inflection was lost in 17 cases (i.e. 42%, which is identical with the percentage for the whole group). Thus, unlike in Dutch possessive pronouns do not appear to trigger or support inflectional loss.

Inflectional generalisation has been mentioned above as a possible path of development, and has been reported for those extraterritorial varieties of Dutch which developed during the 19th and early 20th centuries (East Indian Dutch, Surinamese Dutch and Antillean Dutch, see Van Marle 1995, a preference for inflected forms has also been reported for the Duminy diary, see Scholtz 1963: 129, as well as some Afrikaans dialects, see fn. 18). To investigate the possibility of a (counter-) tendency towards inflectional generalisation involving those monosyllabic adjectives which are today categorically uninflected, only environments where
uninflected forms would have corresponded to the norms of Dutch were considered (N=87), and
20 inflected adjectives were found (23%). At least one individual appears to have adopted
inflectional generalisation as a linguistic strategy: Carolina Leipoldt uses inflected adjectives
almost categorically in constructions involving neuter nouns and indefinite articles (the only
exception is *een groot salaris*).

\[(44) \quad \text{een zoete meisje (2x)}
\]
\[\text{een kleine meisje}
\]
\[\text{een blikke kannetje}
\]
\[\text{een zoete kind}
\]
\[\text{een mooijsje meisje}
\]

It is noticeable that four of these examples involve a head noun which although
grammatically neuter is semantically female (*meisje*), and this could have triggered the inflection
of the adjective. However, in the only instance where Leipoldt uses *meisje* with a definite article
she uses the correct neuter article *het* (see § 6.3.1).

In the data there are 93 examples of polysyllabic adjectives ending in -er (categorically
non-inflecting in Afrikaans) in environments in which inflected forms are required in Dutch, 49
of these adjectives were uninflected (53%). Finally, categorically inflecting monosyllabic
adjectives in the data were analysed by focusing on those cases where according to the system
of Dutch the adjective would have been uninflected, and 57% of these adjectives were found to
be inflected (21 out of 37 cases). The reverse (categorical non-inflection) occurred in only 4%
(14 out of 345 cases). The same procedure was also used for the analysis of polysyllabic
adjectives in the data set, i.e. only cases where the adjective is uninflected in the Dutch system
were included, and 70% of these adjectives were inflected (47 out of 67 cases).

---

28 Cases which involved modes of address (such as 'dear child') were excluded from the analysis
as in Dutch both inflected and uninflected forms (*lief kind* - *liewe kind*) are permissible in such
constructions (see Geerts et al. 1984: 326).
Table 6.15 The emergence of the new system of adjectival declension across the four groups of adjectives (C-1880-1922). (i) number of cases conforming to the new system of Afrikaans, (ii) number of loci where the Dutch system diverges from the new system of Afrikaans (iii) (i) as percentage of (ii) ($\chi^2=20.33505$, df=7, p<0.01).

<table>
<thead>
<tr>
<th>Adj. Type</th>
<th>(i)</th>
<th>(ii)</th>
<th>(iii)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monosyllabic adj. (categorically -o in Afrikaans)</td>
<td>203</td>
<td>479</td>
<td>42%</td>
</tr>
<tr>
<td>Monosyllabic adj. (categorically -e in Afrikaans)</td>
<td>21</td>
<td>37</td>
<td>57%</td>
</tr>
<tr>
<td>Polysyllabic adj. (categorically -e in Afrikaans)</td>
<td>47</td>
<td>67</td>
<td>70%</td>
</tr>
<tr>
<td>Adj. ending in -er (categorically -o in Afrikaans)</td>
<td>49</td>
<td>93</td>
<td>53%</td>
</tr>
</tbody>
</table>

The quantitative data summarised in table 6.14 show that those adjectives which are categorically inflecting in modern Afrikaans conform most closely to the new system. Since there was also a marginal tendency to generalise inflectional -e to those adjectives which today are categorically uninflecting, it is likely that inflectional generalisation which has been reported for other extraterritorial varieties of Dutch (as well as some non-standard varieties of Afrikaans) competed with a complex re-structuring process in which certain subclasses of adjectives proved to be resistant to inflectional marking.

6.4 The Gradualness of Morphological Change

That our understanding of morphosyntactic change is still rather rudimentary was argued by Anderson (1992: 365) who noted that '[u]nfortunately theories of such changes are not well-developed, and real results or established principles are hard to find.' Traditionally analogy (i.e. regularisation of surface forms) has been seen as the main force in morphological change, and Sweet (1900: 23), who described the nature of analogical changes as 'substitution', distinguished the abrupt nature of morphological change from the gradual movement which characterises many phonetic changes. However, although morphological change involves an element of a rather abrupt and discrete process of substitution (or loss) while phonetic changes are often structurally continuous (see, however, Labov 1992: 43-44 on different types of phonetic changes), morphological change is nevertheless gradual and continuous in terms of its temporal (and social) diffusion. In other words, the relative frequency of morphological substitution or loss increases slowly over time. It is exactly this feature of gradual diffusion which makes not only the dating of morphological change difficult, but also the identification of its 'causes'. Thus,
with reference to the history of English, Jespersen (1945: 169) remarked:

This grammatical development and simplification has taken place not suddenly and from one cause, but gradually and from a variety of causes.

For the South African Philological School morphological change at the Cape was gradual, stretching over roughly 200 years (1652-1850), while those emphasising the role of contact in the early settlement have argued for a more abrupt simplification process which was largely completed around 1700, i.e. resulted in the emergence of basilectal varieties which then gradually influenced the linguistic systems spoken in the superstrate (see § 2.2.2). Both schools agree, however, that the morphological simplification process was completed in all varieties by the latest the mid-19th century. Yet, the variable concord phenomena found in the corpus (1880-1922) show that the diffusion of the simplificatory changes which proceeded at different speeds through linguistic environments, was still in process in the more acrolectal varieties around 1900. Summarising the relative frequencies for the different morphosyntactic variables, one sees the following cline for the degree of morphological reduction:

GENDER ≥ PRESENT TENSE > PAST PARTICIPLES (-T) > INFINITIVE > PAST PARTICIPLES (-EN)

The percentages for the corpus are: 55% > 54% > 45% > 40% > 32%, and the frequency differences are statistically significant ($\chi^2 = 295.4275$, df=9, p< 0.001). Note, however, that the marginal difference between loss of gender and loss of person/number marking is not statistically significant. The gradual pattern (including grammatical, phonological as well as possibly some lexical constraints) agrees with Meillet's (1967: 111) dictum on the nature of morphology and morphological change: 'Morphology is the domain of survival.' The distribution patterns presented in this chapter support not only the interpretation of inflectional loss at the Cape as a gradual sequence, but also shows that the 'principle' of inflection (although variable) remained active until fairly late (at least in acrolectal and possibly upper mesolectal varieties of CDV).

To emphasise the gradual nature of morphological change is, however, not to say that language contact played no or only a very minor role in the morphological development of Afrikaans, and as argued in § 2.2.1 the contact situation at the Cape was intense and must have had significant linguistic repercussions. It should also be noted that as regards inflectional loss, both internal West Germanic language change and structural convergence under language contact show the same tendency towards analogical simplification and regularisation, thus, supporting an interpretation in terms of mutual reinforcement leading to accelerated inflectional reduction (§ 2.2.3, see also Roberge's 1998: 31).
MORPHOLEXICAL AND SYNTACTIC VARIATION

Data! data! he cried impatiently, I can't make bricks without clay.
Arthur Conan Doyle The Copper Beeches

7.1 The Pronoun System
7.1.1 Subject and Object Pronouns

The pronoun system of continental Dutch differentiates between subject and object forms, as well as stressed (strong) and unstressed (weak) forms (table 7.1; for a summary of the etymological relations between strong and weak pronouns, see Zwart 1997: 117-118). In Afrikaans the differentiation between subject and object forms was maintained for singular pronouns, but lost in the plural. The unstressed forms were entirely lost (table 7.2; for details see Donaldson 1993: 123ff.). The formal 2nd person pronoun u is omitted from table 7.1 and table 7.2., as well as the (Flemish) 2nd person pronouns gjij/ge which were common in written Early Modern Dutch.

<table>
<thead>
<tr>
<th></th>
<th>subject</th>
<th></th>
<th>object</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>stressed</td>
<td>unstressed</td>
<td></td>
<td>stressed</td>
</tr>
<tr>
<td>1.sg.</td>
<td>ik (ikke)</td>
<td>k</td>
<td>mij</td>
<td>me</td>
</tr>
<tr>
<td>2.sg.</td>
<td>jij</td>
<td>je</td>
<td>jou</td>
<td>je</td>
</tr>
<tr>
<td>3.sg.</td>
<td>hij, zij, het</td>
<td>ie, ze, 't</td>
<td>hem, haar, het</td>
<td>'m, d'r, 't</td>
</tr>
<tr>
<td>1.pl.</td>
<td>wij</td>
<td>we</td>
<td>ons</td>
<td></td>
</tr>
<tr>
<td>2.pl.</td>
<td>jullie</td>
<td>je</td>
<td>jullie</td>
<td>je</td>
</tr>
<tr>
<td>3.pl.</td>
<td>zij</td>
<td>ze</td>
<td>hun (hen)</td>
<td>ze</td>
</tr>
</tbody>
</table>

Table 7.1 The pronoun system of continental Dutch.¹

Raidt (1991: 213) described the loss of unstressed forms in a pragmatic-psychological fashion as the result of the 'neiging om nadruklik te praat'. However, Zwart (1997) has shown that the unstressed Dutch pronouns are not merely phonologically reduced variants of the stressed pronouns, but syntactic clitics which have (i) a specialised meaning that the stressed pronouns lack (i.e. je and ze (pl.) can be used to refer to the generic entity 'people', and ze (pl.) can be used to refer to both persons and things), and (ii) their syntactic status differs from that of the stressed pronouns, i.e. they can move to positions unavailable to stressed pronouns (Zwart 1997: 119ff.). To describe the weak pronouns as special clitics means that their loss involved both lexical and

¹ The general object pronoun for the third person singular is hun. Hen is used in more formal registers as a direct object and after prepositions (Donaldson 1981: 53-54, Geerts et al. 1984: 171).
syntactic change. This section, however, is not concerned with the loss of weak pronouns, but focuses primarily on the development of Afrikaans lexical variants of the strong pronouns.²

<table>
<thead>
<tr>
<th></th>
<th>subject</th>
<th>object</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.sg.</td>
<td>ek</td>
<td>my</td>
</tr>
<tr>
<td>2.sg.</td>
<td>jy</td>
<td>jou</td>
</tr>
<tr>
<td>3.sg.</td>
<td>hy, sy, dit</td>
<td>hom, haar, dit</td>
</tr>
<tr>
<td>1.pl.</td>
<td>ons</td>
<td>ons</td>
</tr>
<tr>
<td>2.pl.</td>
<td>julle</td>
<td>julle</td>
</tr>
<tr>
<td>3.pl.</td>
<td>hulle</td>
<td>hulle</td>
</tr>
</tbody>
</table>

Table 7.2 The pronoun system of Afrikaans.

As for the reduction of the verbal and nominal system the loss of the subject/object distinction has been compared to developments in creole languages which generally do not mark case in the pronoun system (Ponelis 1993: 218). Loss of case distinctions, however, is also a general characteristic of the development of the West Germanic languages where accusative and dative pronouns fell together into one category ('oblique' case) in English and Dutch (but are still differentiated in Yiddish and German; see Sapir 1921: 163 for an interpretation of the levelling of the subject-object distinction as an example of 'drift').³ Nominativisation of object forms has also been reported for dialectal Dutch (see, for example, the use of subjective jou in Stadsfries as reported by De Vooy 1953: 77, or the use of hun in subject position in colloquial Dutch; for further discussion see Scholtz 1963: 98-101, Ponelis 1993: 218). Thus, as in the case of morphological reduction, it is possible that tendencies of internal Germanic language change and contact-induced restructuring coincided and mutually reinforced one another. The following subject and object pronouns were quantified in the corpus:

² It seems that the use of weak pronouns was stylistically constrained in 19th century CDV. Thus, Changuion (1844: 76) writes: 'Je behoort tot de taal des beschaafden omgangs [...] maar jij en jou is plat Holl.' Scholtz (1963: 54-55) noted that the loss of weak pronouns has been described also for other ET varieties of Dutch, and attributes the process to the language use of interlectal speakers who generalised the perceptually salient strong pronouns. Weak pronouns still occur in the acrolectal corpus (C 1880-1922) but are relatively marginal (for example, for the 1st pers. pl. pronoun the ratio between wij - we is roughly 7 to 1). In modern Afrikaans weak pronouns have survived in some lexicalised forms, such as dankie which is derived from dank je, or ditsem which is derived from dit is 'm. According to Ponelis (1993: 196) these lexicalised forms 'might be taken as evidence of the fact that weak pronouns had gained a toehold in early Afrikaans.'

³ In general the development oblique → nominative is not uncommon in IE languages. Note also that the Yiddish and Austrian German/ Bavarian 1st person pl. subject mir is originally a dative, as is the English 2nd person sg./ pl. you (from òow dat/acc. plural; see Lass 1997a: 254). Jespersen (1909: 271) speaks in this context of the 'tendency to let the objective case prevail over the subjective case' existing in 'a great many languages.'
1st person singular subject pronoun (ek): The Afrikaans first person singular pronoun ek still varied in the written language with the Dutch form ik at the beginning of the 20th century (Malherbe 1917: 74f.). Interestingly, spelling variation between i and e rarely affected the first person singular pronoun in texts otherwise characterised by a high frequency of i/e spelling variation. 3652 instances of the 1st person singular form were tabulated, with 858 examples of ek (i.e. 24%). Thus, as already suggested by RAIDt's variationist study of 19th century vernacular literature ([1992]: 303, see also Van Rensburg 1984: 148f.), use of ek was still rare at the turn of the century and most writers continued to (at least) write the Dutch form ik.

3rd person singular subject/object pronouns (dit, hom): Although Afrikaans has lost grammatical gender (§ 6.3.1) anaphoric pronouns still reflect semantic gender differentiation in the third person singular, and semantically male or female antecedents require the use of hy and sy respectively. This distinction has been lost in some non-standard varieties of Afrikaans where the masculine pronoun hy combines with female antecedents (Links 1989: 79f., Van Rensburg 1989: 147, Ponelis 1993: 213), a feature which Roberge (1996: 140) attributed to basilectal influence, and which is absent from the acrolectal corpus.

The loss of the Dutch 3rd person sg. neuter subject pronoun het and the rise of the Afrikaans form dit illustrates the interaction of lexical and syntactic change in the development of the pronoun system. The syntactic position of the Dutch object pronoun het/ 't (3rd sg.) is more restricted than the position of the demonstrative dit which replaced het/ 't in Afrikaans. The normal position for het/ 't as an object pronoun is to the right of the V-cluster and to the left of the indirect object, where it stays even if further transformations are applied to the sentence (Seuren 1997: 230).

(1)    Je hebt het haar gegeven.
       Heb je het haar gegeven?

In some varieties of Dutch movement of the pronoun to the right of the indirect object is possible for unstressed pronouns, thus emphasising their status as special clitics. However, the position next to the verb is generally felt to be more standard (see Zwart 1997: 123-124).
(2) Hij heeft 't 'r gegeven.
Hij heeft 'r 't gegeven.

*Het can occur sentence initially only as a dummy-subject, or as a subject referring to a previous de-noun if the predicate following the verb zijn is a noun (or an adjective used as a noun) preceded by an article.

(3) a. *Het regent.

b. Wie is die menner?
Het is de burgermeester. (Geerts et al. 1984:169)

The Afrikaans neuter pronoun dit is derived from the Dutch demonstrative dit ('this', alongside dat 'that') which can occupy syntactic positions unavailable to het. While het (as direct object) usually criticises to the verb, the demonstratives dit and dat can occur in non-clitic position, are stressable and can be used contrastively (Ponelis 1993: 209f.).

(4) *Het gaf hij mij.  
  Dit gaf hij mij.
  *Ik wil het hebben niet het.  
  Ik wil dit hebben niet dat.

Frequent use of non-clitic het in 18th century documents indicates the breakdown of the Dutch system and has been interpreted as 'masked dit/dat' (Ponelis 1993: 210, also Scholtz 1963: 57-58). Examples of non-clitic het as direct object also exist in the corpus.

(5) neem mij het niet kwaaljk (C. Stanford, 6/8/1905)
hier heeft ook een paar vrouwen het gebruikt (C. Leipoldt 21/8/1880)

In 18th century CDV het competed initially not only with dit but also with the demonstrative dat, and only from the early 19th century dit dominated over dat in replacing the Dutch pronoun het (Raidt 1983: 156, Ponelis 1993: 211-212). The analysis of the quantitative distribution of het, dit and dat in the corpus 1880-1922 (as referential subjects and objects, following Ponelis' quantitative analysis of the Barbier corpus 1993: 210-211) shows that het was still in a strong position in acrolectal registers, competing only with dit, while use of dat was negligible.

<table>
<thead>
<tr>
<th></th>
<th>het</th>
<th>dit</th>
<th>dat</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>1272</td>
<td>793</td>
<td>31</td>
<td>2096</td>
</tr>
<tr>
<td>% of total</td>
<td>60%</td>
<td>39%</td>
<td>1%</td>
<td></td>
</tr>
</tbody>
</table>

Table 7.3 Het, dit and dat as referential subjects and objects (C-1880-1922).

Similarly to variation between ik–ek, variation between hem–hom continued until the
early 20th century (Malherbe 1917: 79). A total of 451 instances of the third person singular masculine object pronoun were tabulated, and 69 instances of *hom* were found (15%).

*1st person plural subject pronoun (ons)*: Early examples of nominative *ons* are prominent in the Khoe-Dutch and Slave-Dutch data (see Scholtz 1963: 93-94), and were reported for the settlers’ varieties of CDV first by Otto Friederich Mentzel who was in the service of the VOC between 1731 and 1733. In the account of his stay, he made the following (frequently cited) remark:

> The language of the country is just as far from being pure Dutch, as that of the German farmers is from pure German. The men have a broad accent and the women folk use certain expressions that are sometimes really ridiculous. For instance, if one were to ask them whether they have no Bible, the reply is: 'Onz heef geen Bybel'; which means 'Us has no Bible'. If one were then to ask them: 'How many 'Onze' (ounces) in a pound?' they would blush. (Mentzel 1944:[1787]): 157-158

This commentary suggests not only that nominative *ons* occurred in the speech of the settlers, but also that speakers recognised the variant as being 'incorrect'. Nominative *ons* occurs in the transcription of an utterance by Hanna Wagenaar in 1772 (CJ 1103/225), which is generally remarkable for its similarity to modern Afrikaans (i.e. first instance of *hulle*, also shows reduction of the verbal system; see Scholtz 1963: 112, Ponelis 1993: 217).

(6)  

> het Hulle dan Overal geloop waar ons geloop het

Scholtz and Raidt have argued that since nominative *ons* is known in older forms of dialectal Dutch, and is still used in the Zeeland dialects, it should be interpreted not as a result of a creolisation process but as a 'Germanic vulgarism' (Raidt 1983: 155, Scholtz 1963: 94-95). However, speakers of Zeeland varieties had little impact on the composition of the early society at the Cape, and since nominative *ons* is not attested in Holland varieties of Dutch which featured prominently in the Dutch base of Afrikaans, direct dialect origin is unlikely (Ponelis 1993: 218). The relatively late attestations of nominative *ons* in the settlers' vernacular contrast with the abundance of early examples of *ons* in the Khoe and slave data (see the data given in § 2.2.1), and Ponelis has, therefore, interpreted Afrikaans *ons* as an 'interlectally restructured form' of Dutch *wij/we* (Ponelis 1993: 32), a 'linguistic innovation among Khoe and slave interlectals' (*ibid.*: 218, see also Valkhoff 1966: 222). While most Afrikaans linguists assume that nominative *ons* was fairly common by 1750 (Raidt 1983: 155, 1991: 213, Ponelis 1993: 218), Hoogenhout (1904: 9) still reported variation between *ons* and *wij* in subject position among older people at the beginning of the 20th century.

1236 examples of the 1st person plural pronoun in subject position were tabulated and
"ons" occurred in 448 cases (i.e. 36%). Its distribution across individuals (in percentages) is summarised in figure 7.1.

![Histogram](image)

**Figure 7.1** Histogram for 65 cases, summarising the use of nominative "ons" across cases (based on percentages for individuals).

The distribution of cases is again polarised, with conformity to the norms of Dutch being more pronounced than in the distributions for loss of gender and the present tense paradigm. Noticeable is the small and discontinuous middle field where only 28% of the individuals are found (most of them in the area towards the left, i.e. showing frequencies <50%).

Individuals who do not (or rarely) use nominative "ons" might nevertheless show considerable reduction in the verbal system. In other words, although use of "ons" generally triggers the uninflected form of the verb, "wy" nevertheless occurs frequently with uninflected verbs in the corpus. Note that there also seems to have been a tendency towards co-variation of pronoun variants in the speech of individuals:

De enige Afrikaners, die *ik* in plaats van *ek* in een gewoon gesprek gebruiken, bezigen *wy* of *wy is* in plaats van *ons* is. (De Goede Hoop, November 1906, 111; see also the analysis in §8.1.1)

3rd person plural subject/object pronouns (*hulle*): The third person plural subject pronoun *hulle* is usually interpreted as a variant of the Dutch object form *hunlui"/"hunlie(den)*, and shows thus the same grammatical process as *ons* (nominativisation). *Hulle* competed as a subject form with *zulle/sulle* (derived from *sylui*) during the 18th century. It appears that *zulle/sulle* was considered more 'genteel' (Ponelis 1993: 219), and it occurs, for example, in Johanna Duminy's diary (see...
§ 2.1) in variation with the standard forms (zij/sij and haarlui). Use of nominative hulle increases after 1810 (Scholtz 1963: 96); however, variation with zulle continued until the early 20th centuries (Malherbe 1917: 80). Hulle as an object form competed with haarle (with the variants haarly and haarlui) which was the dominant form throughout the 18th and early 19th century, and use of objective hulle increased only after 1840 (Scholtz 1963: 97). Like z/sulle, haarle was still used by older people at the beginning of the 20th century (Malherbe 1917: 80).

In the corpus hulle (with the variants hul and helle) is widespread in subject position (49%, i.e. 172 out of a total of 352 instances), but rarer in object position (32%, i.e. 51 out of 160). The greater prominence of hulle in subject position agrees with the patterns of variation reported for 18th and early 19th century CDV (see Scholtz 1963: 99, Ponelis 1993: 220).

<table>
<thead>
<tr>
<th></th>
<th>subject</th>
<th>object</th>
</tr>
</thead>
<tbody>
<tr>
<td>hulle</td>
<td>172</td>
<td>51</td>
</tr>
<tr>
<td>N</td>
<td>352</td>
<td>160</td>
</tr>
</tbody>
</table>

hulle as
% of total 49% 32%

Table 7.4 Hulle as subject and object pronoun (C-1880-1922). \( \chi^2 = 13.34180, \text{df} = 3, p < 0.004. \)

The variant zulle/sulle is rare in the corpus and only thirteen instances were found. Interesting is the use of zulle in a letter by J.H. De Villiers to S.J. Du Toit (22/4/1891). At the beginning of the letter De Villiers announced as his aim to write in the kind of language he uses in ordinary conversation ('Ik wil nou probeer om in myn brief dieselfde taal te gebruik als ik in gewone gesprek zou gebruik'), and on the whole the language used in the letter is close to the emerging standard of Afrikaans. Yet, his only example of a 3rd person plural pronoun (subject) is zulle, and he, furthermore, criticises Du Toit's use of nominative ons.

Bij voorbeeld, u spel altyd 'veel' as 'veul' en zeg 'ons' waar in gewone gesprek 'wy' gezegd wordt.

In sum, the frequency patterns for the different pronoun variables in the corpus are in line with the results from other analyses, that is, very low frequencies were found for ek and hom, the 3rd person sg. neuter variant dit was clearly dominant over dat but still in competition with het, and the 3rd person pl. form hulle was more prominent in subject than in object position. The results for the six pronoun variables are summarised in table 7.5.
<table>
<thead>
<tr>
<th></th>
<th>ek</th>
<th>dit</th>
<th>hom</th>
<th>ons</th>
<th>hulle (s)</th>
<th>hulle (o)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>858</td>
<td>793</td>
<td>69</td>
<td>448</td>
<td>172</td>
<td>51</td>
</tr>
<tr>
<td>Total</td>
<td>3652</td>
<td>2096</td>
<td>451</td>
<td>1236</td>
<td>352</td>
<td>160</td>
</tr>
<tr>
<td>% of total</td>
<td>24%</td>
<td>39%</td>
<td>15%</td>
<td>36%</td>
<td>49%</td>
<td>32%</td>
</tr>
</tbody>
</table>

Table 7.5 Summary: The subject/object pronoun variables (C-1880-1922). \( \chi^2=263.6439, \text{df}=11, p<0.001 \).

### 7.1.2 Attributive Possessive Pronouns

The Afrikaans attributive possessive pronouns are (except for the third person singular masculine/neuter) identical with the object forms of the personal pronouns (see Geerts et al. 1984: 201-205 for general overview over the Dutch system, Donaldson 1993: 123, 138f. for Afrikaans).

<table>
<thead>
<tr>
<th></th>
<th>'stressed'</th>
<th>unstressed</th>
<th>Afrikaans</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.sg.</td>
<td>mij</td>
<td>m'n, me</td>
<td>my</td>
</tr>
<tr>
<td>2.sg.</td>
<td>jouw</td>
<td>je</td>
<td>jou</td>
</tr>
<tr>
<td>3.sg.</td>
<td>zijn, haar</td>
<td>z'n, d'r</td>
<td>sy, haar</td>
</tr>
<tr>
<td>1.pl</td>
<td>ons/once</td>
<td>ons</td>
<td></td>
</tr>
<tr>
<td>2.pl</td>
<td>jullie</td>
<td>julle</td>
<td></td>
</tr>
<tr>
<td>3.pl</td>
<td>hun</td>
<td>'r, d'r</td>
<td>hulle</td>
</tr>
</tbody>
</table>

Table 7.6 Attribute possessive pronouns in Dutch and Afrikaans.

There were important 'crossovers between object and possessive pronouns' in early Dutch and some Dutch dialects, which affected, amongst others, the first person singular pronoun, where variation between mij and mijn has been reported for the object pronoun, as well as the attributive possessive pronoun (Ponelis 1993 204, also De Vooy 1953: 75). The archival data for CDV show that at the Cape the n-form dominated in both functions during the early 18th century.

---

4 Variation between objective mij and mijn was a social marker in early modern Dutch. The Statenvertalers working on the Dutch version of the Bible rejected the variant mijn as too low and colloquial: 'unquam myn, ut vulgus hic loquitur', i.e. 'never use mijn as lower class people do' (Van Der Wal 1992: 124).
<table>
<thead>
<tr>
<th></th>
<th>object</th>
<th>possessive</th>
<th>both</th>
</tr>
</thead>
<tbody>
<tr>
<td>mijn</td>
<td>74</td>
<td>46</td>
<td>120</td>
</tr>
<tr>
<td>mij</td>
<td>50</td>
<td>31</td>
<td>81</td>
</tr>
<tr>
<td>total</td>
<td>124</td>
<td>77</td>
<td>201</td>
</tr>
<tr>
<td>mijn as % of total</td>
<td>60%</td>
<td>60%</td>
<td>60%</td>
</tr>
</tbody>
</table>

Table 7.7 *Mijn*/*mij* in Early Afrikaans 1710-1750 (from Ponelis 1993: 205).

An even clearer preference for *mijn* (in both functions) existed during the late 18th and early 19th centuries, but from around 1840 forms without -*n* came to dominate in object position (see the data given in Ponelis 1993: 205), while *mijn* remained an important variant of the attributive possessive until the early 20th century (*ibid.*: 226-7, also Malherbe 1917: 82).5

Objective *mijn* occurs in the corpus, but is rare (6%, i.e. 58 out of a total of 1054 object forms), and limited to a small group of speakers (all older generation, age group 1) who have generalised it (B.J. Brümmer, D.J. LeRoux, and H.H. Pienaar). With regard to the possessive pronoun, the -*n* form is still widespread, and possessive *mij* occurs in only 28% of cases (262 out of 940). The percentage is lower for *zij/sij* as a possessive pronoun, a tendency also noticeable, for example, in the vernacular texts of Zwaartman where the respective percentages are 60% for *mij* and 49% for *sij* (Ueckerman 1987: 15, Ponelis 1993: 227). In the plural, use of possessive *hulle* (3rd person pl.) is relatively widespread in the corpus: 38 out of 94 cases (i.e. 40%).

<table>
<thead>
<tr>
<th>mij</th>
<th>zij/sij</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>262</td>
</tr>
<tr>
<td>Total</td>
<td>940</td>
</tr>
</tbody>
</table>

*Mijn*- *zij/sij* as % of total 28% 15%

Table 7.8 *Mijn* and *zij/sij* as possessive pronouns (C-1880-1922). $\chi^2=28.51365$, df=3, p<0.001.

In Dutch the possessive pronoun for the 1st pers. pl. concords with the head noun: *ons* is used before neuter sg. nouns (*ons* boek), *onze* before all other nouns (*onze* hond, *onze* boeken) (Geerts *et al.* 1984: 201-202). Although Afrikaans has lost nominal gender and has generalised the form *ons*, variation between *ons* and *onze/onse* was reported by Malherbe as typical for the

5 While *mij* dominates in object position, for example, in the writings of Meurant, Boniface and Zwaartman, the GRA publications *Die Patriot* and *Ons Klyntij* show a clear preference for *mijn* (73%). Preference for objective *mijn* is also evident in Abu Baki's *Bayānudīn* (1869, see Van Selms 1979, Ponelis 1993: 205, see also Lätti 1978: 114). Variation continued until the early 20th century but was evaluated negatively. Thus Hoogenhout (1904: 9) remarked in his grammar: 'Akkusative Singular *myn* für *my* ist nicht nachzuahmen.'
OFS and the Transvaal, as well as for formal registers which aimed at a 'Nederlands gekleurde taalgebruik' (Malherbe 1917: 82-83). Onse has also been maintained in some varieties of Western Cape Afrikaans (Ponelis 1993: 216), and its seemingly archaic and formal nature can be exploited for ironic and humorous purposes (Donaldson 1993: 140). In the corpus 275 instances of ons and 139 instances of onze/onse were found (table 7.9).

<table>
<thead>
<tr>
<th>neuter singular nouns</th>
<th>common gender and plural nouns</th>
</tr>
</thead>
<tbody>
<tr>
<td>ons</td>
<td>onse</td>
</tr>
<tr>
<td>82</td>
<td>193</td>
</tr>
<tr>
<td>14%</td>
<td>61%</td>
</tr>
<tr>
<td>% of total</td>
<td>39%</td>
</tr>
</tbody>
</table>

Table 7.9 Ons/Onze as attributive possessives (C-1880-1922). $\chi^2 = 62.70755$, df=3, $p < 0.001$.

The frequencies summarised in table 7.9 indicate that ons was in the process of being generalised to common singular and plural nouns, the majority of which already combine with ons. At the same time, onze was found to combine marginally with neuter singular nouns.

7.1.3 The Relativiser

In continental Dutch the form of the relative pronoun in direct relative clauses is determined by the gender of its NP antecedent: die is used for common gender nouns and plurals, dat (and its colloquial and dialectal variant wat) for neuter singular nouns. In standard Dutch the form wat occurs regularly after some indefinite pronouns (dat is iets wat ik al wist) and in clause-linked relatives for which no direct antecedent exists (as in ik kan niet krijgen wat ik nodig heb). In Afrikaans only a single relativiser (wat) is used (for Dutch see Donaldson 1993: 63-65, Geerts et al. 1984: 245-252, for Afrikaans Donaldson 1993: 145-149).

(7) Dt. de man die Piet heet het meisie dat Engels praat de meisies die Engels praten
     Afr. die man wat Piet heet die meisie wat Engels praat die meisies wat Engels praten

Middle Dutch did not have a separate class of relative pronouns, but used in the typical West-Germanic fashion demonstratives (D-forms, die, dat, daar, diens) and interrogatives (W-forms, wie, wat, waar, wiens, welke; Burridge 1992: 243). In Middle Dutch interrogative pronouns occurred mainly in prepositional relatives (as is also the case in standard Dutch: de vrouw met wie ik gisteren praatte), and soon dominated after indefinite pronouns, although D-forms still occurred (Ponelis 1986: 53-54). Use of wat with neuter singular antecedents was not

---

6 Malherbe also mentions the possibility that onse might have been re-analysed as ons se, i.e. as a periphrastic possessive (Malherbe 1917: 83, see § 7.3.4).
uncommon in 19th century Dutch, and is also attested for dialectal varieties of Dutch where it competes with the D-form *dat (Sassen 1983, Geerts et al. 1984: 250-251). Use of W-forms (interrogatives) as relativisers is widespread in other West Germanic languages (English, Yiddish, southern German dialects, Frisian as well as the Dutch creoles Negerhollands and Berbice Dutch; see Kouwenberg 1994: 236, Den Besten 1996: 16-17).

Several linguists have argued for the need to distinguish between relative pronouns and relative markers (or complementizers; see the overview in Lightfoot 1979: 314, Romaine 1982a: 58, Den Besten 1996). Unlike relative pronouns, relative markers are invariant (i.e. they do not vary according to the coreferential NP), cannot serve as prepositional objects, and lead to preposition stranding. English *that, for example, is not a pronoun but a complementizer: *this is the university at that she works (but periphrasis is possible: this is the university [that] she works at). Similarly Frisian *wat, Bavarian *vos, Yiddish *vos, Berbice Dutch *wati and Afrikaans *wat might not warrant classification as pronouns, and are probably better described as relative markers or complementizers (as argued in detail by Den Besten 1996). Thus, Afrikaans *wat is not only invariant, but also cannot be the object of a preposition. Afrikaans has maintained the (Dutch) constructions *P+ *wie (with human antecedent) and *waar + P (with non-personal antecedent) to introduce prepositional relatives. However, in cases of preposition stranding *wat is obligatory (Den Besten 1996: 12).

(8) die vrouw met wie ik gister gepraat het
die problem waarvan jy gepraat het
die vrouw wat ek gister mee gepraat het
die problem wat jy van gepraat het

That the use of an invariant relative marker is not dependent on loss of gender can be seen from the situation in Yiddish which has maintained a nominal three gender system, but uses the invariant relative marker vos in addition to the inflected relative pronoun velkh (Lowenstamm 1977, Weinreich 1992 [1949]: 197, 306, Jacobs, Prince & Van der Auwera 1994: 416; the same is true for some High German dialects as well as Frisian). 7

7 There is a good argument for Yiddish vos being a complementizer in cases where it refers to a human antecedent. VOS cannot be the object of a preposition unless the clause contains a resumptive pronoun which then functions as the head: * der mensh mit vos ikh hob geredt, but der mensh vos mit im hob ikh geredt; or unless the antecedent is nonhuman *di teme af vos er redi iz ... (Lowenstamm 1977, Jacobs, Prince & Van der Auwera 1994: 416, Weinreich 1992[1949]: 198). Invariant relativisers are widespread in North Germanic: Danish, Icelandic, Norwegian (bokmål) and Swedish. An invariant relative marker is also used in Pennsylvania German (as 'that'; see the respective articles in König & Van der Auwera 1994).
(9) Der mentsh vos zitst baym tish iz mayn tate.
    Dos bukh vos ligt oyfn tish iz interesant.

However, loss of gender probably facilitated the generalisation of *wat* to common gender and plural antecedents, as argued by Ponelis (1993: 193):

> [G]ender helps to delimit the respective domain of D- and W-relatives (as common gender *die* vs. neuter *wat* in current Dutch), and neutralisation of gender, as in English, Negerhollands and Afrikaans, does remove an obstacle in the path of W-relatives.

Ponelis (1987: 64-65, 1984) suggested that generalisation of the neuter relativiser to common gender and plural antecedents might have been influenced by colonial Dutch usage which shows occasional examples of *dat* (but not *wat*) as an invariant relative. However, Den Besten (1996: 23) has shown that the use of an invariant relativiser, although widespread in Germanic, is not attested for Holland varieties of Dutch, which featured prominently in the dialectal basis of Afrikaans (see § 2.2.1). In other words, while the development of an invariant relative marker is widespread within the language family, direct dialect origin is unlikely. Within the substrate an invariant relative marker is attested for Low Portuguese (*que*) (Ponelis 1993: 193), and Hesseling (1923: 121) has suggested that influence from Malay, which uses the invariant relative marker *yang* could also have played a role. However, since the Malay relative construction differs structurally from the relative construction in Afrikaans, influence was possibly minor (Ponelis *ibid.*).

425 relative clauses with lexical NP antecedents were tabulated (six relative clauses which made use of the formal pronoun *welk(e)* were excluded). Although the corpus contains several instances of relative *wat* following a common gender antecedent (105 out of 336 i.e. 31%), the opposite strategy, that is, use of relative *die* following neuter singular NPs, is dominant (46% of neuter singular NPs are followed by *die*, i.e. 41 out of 89). Generalisation of the common gender and plural relative *die* to neuter singular NPs parallels the development in the gender system where the common gender and plural article *de/die* was generalised to neuter singular nouns (see § 6.3.1). Note that in those cases where a *die*-relative refers to a neuter singular noun, the antecedent NP is either indefinite or loss of gender is already noticeable in the determiner.

(10) die Geld die wij voor Mr Hendrickse gecollecterd heeft (Aletta April, 27/1/1913)
     een jong meisje die naar school gaat (Katie van Huyssteen, May 1907)

---

8 Only noun-phrase linked direct pronominal relatives (restrictive and non-restrictive) were considered, while prepositional relatives, nonpronominal relatives and possessive relatives were excluded from the quantitative analysis.
Generalisation of relative *die* was common in 19th century CDV, and variation between *die* and *wat* has been described for Trigardt's diary (Smuts 1968) as well as for the vernacular writing tradition (Van Rensburg & Combrink 1984, Raidt 1994 [1992], see § 3.1). It is possible that the choice between *die* and *wat* was somehow stylistically motivated. This is suggested by the discussions surrounding the Bible translation, where *wat* was felt by many not to be appropriate when referring to God (see Steyn 1989: 26, see also NGK archives, AS 15/3).

In sum, in the corpus about one third of relative clauses (32%) use *wat* as a relativiser. The difference between common and neuter gender antecedents is small (31% for common gender antecedents, and 34% for neuter gender antecedents), suggesting that loss of gender probably occurred prior to the generalisation of *wat*. The relatively high percentage of *wat* contrasts with the 18th century CDV data analysed by Ponelis (1993: 193-194), where its use was still rare. Although *wat* is well established in the corpus, generalisation of *die* as an invariant relativiser was clearly the dominant strategy; use of the neuter D-form *dat* is marginal.

<table>
<thead>
<tr>
<th></th>
<th><em>die</em></th>
<th><em>dat</em></th>
<th><em>wat</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Common gender/ plural antecedent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>225</td>
<td>6</td>
<td>105</td>
</tr>
<tr>
<td>% of total</td>
<td>67%</td>
<td>2%</td>
<td>31%</td>
</tr>
<tr>
<td>b. Neuter gender antecedent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>41</td>
<td>18</td>
<td>30</td>
</tr>
<tr>
<td>% of total</td>
<td>46%</td>
<td>20%</td>
<td>34%</td>
</tr>
<tr>
<td>c. Both</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>266</td>
<td>24</td>
<td>135</td>
</tr>
<tr>
<td>% of total</td>
<td>63%</td>
<td>6%</td>
<td>32%</td>
</tr>
</tbody>
</table>

Table 7.10 The use of relative pronouns following lexical NP antecedents (C-1880-1922).

The conventional dating of invariant *wat* to circa 1775 (Raidt 1983: 160, Ponelis 1987: 68a) appears rather early, when considering the distribution patterns in table 7.10. Den Besten (1996) has argued that the prominence of generalised *die* in the historical record suggests the existence of two competing systems until the late 19th and possibly early 20th centuries, i.e. a simplified Dutch system which made use of *die* as a general relative pronoun as well as of *waar/wie* with prepositional relatives (acro- and mesolectal varieties), and a substratum system using the invariant marker *wat*. From about 1800 use of *wat* increased slowly, but *die* remained part of many varieties of CDV.

---

9 Substrate systems also made use of zero marking, since in Khoe the relative pronoun is sentence final (Carla Luijks, p.c. April 1999).
7.1.4 The Demonstrative Pronouns 'hierdie' and 'daardie'

Modern standard Afrikaans has the demonstratives *hierdie, daardie*, emphatic *dié* and the independent demonstrative *dit* (Donaldson 1993: 142-145). In colloquial Afrikaans a third form *doerdie* ('yonder') is often used in contrast with *daardie* (Ponelis 1993: 169).

(11)  

hierdie/daardie boek is interessant  
dié boek is interessant  
Het *jy dit* gelees? / Het *jy hierdie* gelees?

During the 18th century the Dutch demonstratives *dezeldie* (preceding common gender singular and plural NPs) and *dai/dit* (preceding neuter singular NPs) were still common in CDV (Scholtz 1963: 126-128). There is but one single case of *daardie* for the 18th century in a letter written by field cornet Floris Visser (1797). Visser reports an incident between a white colonist (*dese Scholtz*) and a black chief (*dadi Kaffer Capiteijn*). In the context of the letter the use of *dadi* is clearly pejorative, while *dese* appears to have been the unmarked form (Raidt 1993: 287-88). Examples of *hierdie* and *daardie* are rare even in second half of the 19th century, and only one example (12) has been found so far for the 1860s.

(12)  

dan zal *dai diedag* bloet geeloopt het (*Het Volksblad* 14.01.1864, quoted in Raidt 1993: 288)^10^  

The forms *hierdie* and *daardie* appear suddenly in the sources from around 1880 without any noticeable period of transition, and are reported in most grammatical and lexicographical descriptions from the late 19th century as the common forms of the demonstrative (see Van der Merwe 1971). To explain the absence of *hierdie/daardie* in the sources with the familiar argument of stigmatisation leading to conscious avoidance (Ponelis 1993: 169) is problematic, since this should not have affected the vernacular writing tradition, which consciously aimed at representing the colloquial language, and did not generally avoid linguistic stereotypes and stigmatised expressions (see § 3.1).

Regarding the origin of the demonstratives, Roberge (1994b: 71) has argued that they go back to basilectal varieties of CDV which used deictic adverbs (*hier, daar*, and *doør*) to specify proximity and distance. In contemporary Griqua Afrikaans the deictic adverb alone can fulfill the function of a determiner (see the examples given in Roberge 1994b: 73), which possibly parallels the situation of an earlier basilectal variety (assuming that Griqua Afrikaans can be understood as a conservative dialect which has maintained earlier non-standard structures; see

---

^10^ Ponelis has, however, pointed out that the phrase is ambiguous since *dai* can also be read as an R-subject (strong form of Dutch *er*) followed by *die dag* (1993: 169).
§ 2.1). Raidt (1993), on the other hand, has interpreted the formation of demonstratives consisting of a deictic locative + definite article as a general Germanic feature since similar constructions have been described for German and Flemish dialects, as well as for Yiddish, English and Swedish.\textsuperscript{11} All these constructions are typical (even as standard constructions) for the spoken language.

\begin{center}
\begin{tabular}{ll}
(13) & German \hspace{1cm} hier das Buch ist sehr interessant \\
& (both standard and dialectal) \\
& Yiddish \hspace{1cm} of der tsimer is groys \\
& English \hspace{1cm} these here people are French \\
& (dialectal) \\
& Swedish \hspace{1cm} den där/här mannen\textsuperscript{12}
\end{tabular}
\end{center}

The use of deictic adverbs to emphasise the demonstrative meaning of the definite determiner is thus not uncommon in Germanic, and has been grammaticalised in Yiddish, Swedish and Afrikaans. This, however, does not exclude influence from basilectal varieties along the lines suggested by Roberge, and mutual reinforcement of the two processes is again possible.

To account for the late occurrence of \textit{hierdie} and \textit{daardie} in the sources, Raidt (1993) has suggested interpreting the situation in terms of Henn-Memmesheimer’s (1986) notion of ‘habitualised’ speech patterns, i.e. although these forms occurred only infrequently, they nevertheless constituted ‘habitualised’ forms which existed side by side with the more frequent forms, especially emphatic \textit{dié}.\textsuperscript{13}


\textsuperscript{12} In Swedish the demonstratives \textit{den/det} preceed the noun and are identical with the front (adjectival) article (as in \textit{den gamle mannen} - the old man), while the definite article for nouns is enclitic (-en/-et).

\textsuperscript{13} Henn-Memmesheimer's approach to linguistic variation is strictly non-quantitative: Es geht darum generalisierte Regelformulierungen zu finden, die jedes belegte habitualisierte Muster erfassen, unabhängig von der Auftretenshäufigkeit. Im Gegensatz zu Leska 1966, Rosenkranz 1970, Stürnemann 1980 geht diese Arbeit nicht davon aus, daß sich kodifizierter Standard und Nonstandard im wesentlichen durch die Häufigkeit einzelner Muster unterscheiden, sondern will zeigen, daß es habitualisierte Muster gibt, die mit Standardgrammatiken nicht beschreibbar sind. In dieser Zielsetzung spielen weder pragmatische Verwendungsbedingungen eine Rolle noch ein sozialer Status, der der Verwendung eines Musters assoziiert ist, sondern \textit{relevant ist ausschließlich die Tatsache, daß ein Muster als 'Sprachgebrauch, als usuell belegbar gilt} (1986: 11) [my emphasis].
Our data indicate that the demonstrative dië was by far the strongest form before 1800, but that there were other, non-standard variants, i.e. the locative reinforced hierdie and daardie which had become habitualised patterns. [...] The meteoric rise of the new forms [after 1880-A.D.] [...] reveals the potential vitality of unnoticed variants in the process of linguistic change. (Raidt 1993: 291)

However, while this interpretation is more persuasive than the customary assumption of stigmatisation, it tells us little about how and why these habitualised but rare variants suddenly assumed generality, and were selected into the emerging standard language. Note in this context that Hoogenhout (1904: 10) described in his Lehrbuch a rather unstable system of demonstratives, and reported the existence of several deictic demonstratives, most of which were not selected into the system of standard Afrikaans: di, hier-di, di-een, deus-kant-sy (all glossed as 'this'), daar-di, di-anner-een, anner-kant-sy (all glossed as 'that'). Only few examples of hierdie and daardie occur in the corpus: there are nine examples of hierdie, and only six examples of daardie.

(14) en denk hier di roos of bloem het mama geplant (B.J. Brümmer, 19/9/1907)
    hierie selfde broer van my is sekretaris (Grosskopf, 1/2/1907)
    om hierie brief te pos (Grosskopf, 26/3/1907)
    ons het hier di laaste tyd vrot weer gehad (Grosskopf 3/7/1907)
    sover als ek kan sien sal ek hierdie week sewe volle ure hé (Grosskopf, 8/12/1907)
    hierie mej. Hofmeyer en haar maat (Grosskopf, 18/9/1909)
    zy heeft aperkatie vir hier die school gemaak (E.Eksteen, 1/1/1922)
    hier die kaartje is van wilde beesten (J.Malan, 1910)
    hier die plek is van die bezigheidplek van Pretoria (J.Malan, 1910)

(15) Jan Krieg [...] die zeg daardie vuiles, en daar die vuiles is julle voor mannen (Schabort, 1902)
    die een wat julle daarie slag op Bloemfontein gesien het (Grosskopf 1/12/1907)
    myn smaak ook [...] in daardie rigting gelop het (Jan de Waal, 3/6/1910)

As for the verbal system, the loss of gender, and the nominativisation of object pronouns, the development of hierdie and daardie is probably best understood as the result of superstrate as well as substrate inputs, i.e. an example of Roberge's concept of 'multilevel syncretism, in which phonological, syntactic and semantic properties of morphemes can be traced to multiple sources' (Roberge 1994b: 73, 1998: 27).

7.2 Some Other Morpholexical Variables ('het', 'baie' and the complementizer 'lat/laat')

The Verbal Variant 'het': The distribution of the third person singular form het (instead of heef(t)) shows a pattern similar to the one described for the pronoun variant ons. That is although some variation exists in the middle field between categorical use and non-use, the overall pattern of variation is bisystemic (846 examples of het and 626 examples of heef(t) were tabulated).
Figure 7.2 Histogram for 70 cases, summarising the use of het across cases (based on percentages for individuals)

Baie: The Afrikaans quantifier baie (derived from Malay banja(k)) has largely displaced the Dutch forms veel and zeer in Afrikaans, and belongs today to the 'top fifty most frequently used Afrikaans lexemes' (Ponelis 1993: 100). The first attestations of baie date from the late 18th century (Scholtz 1965: 153; 1972: 113) and in the 19th century several variants existed: bajang, baing, banjang, banje, banja, baijan, bajan, baain and baija (Smith 1965, Raidt 1994 [1992]: 301). The dominant form in the corpus is the variant baing and 299 examples of the indefinite numeral were found (vs. 353 instances of veel, and 262 instances of zeer).

The Complementizer 'lat/laat': In non-standard varieties of Afrikaans the form lat/laat varies with the standard complementizer dat. Variation between /d/ and /l/ has since LeRoux (1910: 107-108) been interpreted as a result of phonological change, a view which has been questioned by Paardekooper (1990) who suggested Dutch dialect origin, and by Roberge (1994b: 82-84), who has pointed to the existence of the co-variant laat which has so far attracted little scholarly attention. According to Roberge the development of the complementizer lat/laat can be interpreted as a re-analysis of the linking verb laten/laat into a serial verb in the most basilectal varieties, similar to serial 'say' meaning 'that' in Caribbean creoles (Roberge 1994b: 84, see Holm 1988/89 1: 185f.). 14 Only five examples of the form lat/laat as complementizer were found in the data. Use of lat/laat is typical for Malay Afrikaans and ORA (Rademeyer 1938: 53), and three

14 The homophony of the verb laat and the non-standard complementizer lat/laat can lead to hypercorrections in which the verb laat is replaced by the form dat. This is typical after moenie: Moenie dat (=laat) ek jou steur nie (Ponelis 1979: 243).
out of the five examples in the corpus come from Coloured speakers (Cornelius Coopman and Piet Dampies).

(16) toen had die kerkraad besluit laat dit sal voor die gemeente gedrag word (C. Coopman, 18/8/1909)
daarop had die kerkraad dit toegezien laat die onkoste te veel is om aan te sluit aan die senoodes (C. Coopman 18/8/1909)
Meneer Celliers heef gezeg laat de zending commeeze geschryf het (P. Dampies, 30/8/ 1891)

7.3 Selected Syntactic Variables
7.3.1 The Double or Brace Negation

To express negation of a sentence or sentence constituent in Afrikaans a negator (or negative operator, Nie-1) and an obligatory scope marker (or associated negative, Nie-2) are used. The syntactic position of the negator is the same as that of the Dutch negator niet, i.e. within the verb phrase and not post-subject as in many pidgin/creole languages (Ponelis 1993: 461-462). The main negator is the negative particle nie, but as in Dutch negative adverbs can also fill this position. The second negative element (Nie-2) is clause final, and only the lexical form nie can be used. Use of Nie-2 is optional after clause-final negative adverbs, while its omission after clause-final Nie-1 is obligatory. Colloquial varieties of Afrikaans often assimilate Nie-1 and Nie-2 onto the preceding element and use the reduced enclitic form -ie (Donaldson 1993: 401-403, Ponelis 1993: 453ff., for generative account see Wafer 1978, Robbers 1992, Oosthuizen 1998).

(17) ek praat nie graag nie
    ek hoor niemand (nie)
    ek ken hom nie
    ek wil-le slaap-ie

The two negative elements are usually interpreted as scope-markers which can negate entire sentences (simplex and complex sentences) as well as sentence constituents (Wafer 1978: 57, Ponelis 1993: 455-57; for an argument against their scope-marking function see Robbers

---

15 The conventional term 'double negation' (dubbele ontkennning) is not particularly well chosen as it can be seen as including cases of Germanic negative spread (as, for example, possible in German, see Luther's Bible translation 'ein Füllen [...] auf welchem nie kein Mensch gesessen ist' Mark 11:2). The term 'brace' or 'embracing' negation reflects more accurately the phenomenon under description (see Burridge 1992: 178 fn., also Markey 1982: 199).
1997: 39-41). \footnote{Note in this context that in Afrikaans (syntactic) reduplication can also affect particles such as \textit{al} (\textit{ek is al moeg al}) and prepositions such as \textit{in} (\textit{hy woon in hierdie huis in}), however, such constructions are facultative and considered colloquial (see Boonzaier 1989: 282-283 for examples).}

\footnote{Leftward movement of the modal auxiliary is still common in subordinate clauses in ORA, while in Standard Afrikaans the modal auxiliary occupies the second last position (Webb 1993: 165). According to Webb a similar tendency is common in the Afrikaans of White speakers in Pretoria (ibid.)}

(18) a. sentence negation
Wie besef dit \textit{nie} goed \textit{nie}.
Wie besef dit \textit{nie} goed dat die stelsel verander moet word \textit{nie}. (Ponelis 1993: 456-457)

b. constituent negation
\textit{Nie} die man \textit{nie} maar die vrou was die skuldige (Le Roux 1923: 167)
\textit{Nie} Jan het die werk gedoen \textit{nie}, maar Piet. (Robbers 1992: 231)

Regarding the origin of the double negation both dialectal Dutch influence and language contact (substrate influence) have been advanced as explanations. Following Pauwels (1958) Dutch dialect influence has been emphasised by the South African Philological School. Scholtz and Raidt have argued that the Afrikaans negation originates from a facultative Dutch dialect pattern which grammaticalised at the Cape (see Scholtz 1980: 90-91, Raidt 1983: 190, 1991: 223). However, Den Besten (1985/1986) has shown that the syntactic structure of the Afrikaans negation differs significantly from the negation used in dialectal Dutch. That is, in those Dutch (and German) dialects for which double negation patterns have been reported, \textit{Nie}-2 typically occurs clause internally and is thus clearly a phenomenon of the 'middle field'. \textit{Nie}-2 can occur sentence final only if the V2-rule has removed the verb from the final position, or if the internal negator is situated close to the end of the clause. Sentence final \textit{Nie}-2 at the end of complex embedded clauses, such as in example (18a), is not possible in dialectal Dutch. Furthermore, those Dutch dialects for which frequent use of a facultative double negation has been reported, are southern dialects of Dutch (especially Brabant and Flanders), which have been marginal to the development of Afrikaans (see §2.2.1). However, general Germanic negative spread which is clause internal and occurs in a wide variety of dialects of Dutch and German (as well as English), was part of the superstrate input (Ponelis 1993: 467).

Waaler (1994a), who accepts den Besten's argument about the impossibility of a direct continuation of Dutch dialectal negation patterns, nevertheless maintains that Dutch dialect influence was important for the development of the Afrikaans negation. She argues that there was a general tendency in 18th century CDV to move modal auxiliaries, adverbials as well as the negation particle to the left of the verbal complement, often leading to the replication of the moved element in the original position.\footnote{Leftward movement of the modal auxiliary is still common in subordinate clauses in ORA, while in Standard Afrikaans the modal auxiliary occupies the second last position (Webb 1993: 165). According to Webb a similar tendency is common in the Afrikaans of White speakers in Pretoria (ibid.)}
The leftward movement of the negator and the resulting replication of the negative particle in the original position led to a situation in which a second sentence internal negative became common in CDV, where it possibly functioned initially as an emphasis marker. This development was supported by existing Dutch dialectal patterns which allowed for the placement of a second negative in the middle field.

In a second step the sentence internal second negative was increasingly pushed to the right, i.e. became sentence final, lost its function as an emphasis marker and grammaticalised into a scope marker. The transition period of this development was, according to Wafer, characterised by the occurrence of multiple sentence internal negative particles. Such triple (and even quadruple) negation patterns remained productive until the early 20th century (Van Oordt 1913: 83), and are still common in colloquial Afrikaans (see Feinauer 1994, Rademeyer reports such constructions for Griqua Afrikaans 1938: 74, and Boonzaier 1989: 281 for Western Cape Afrikaans).

Of was dat, tusse de voornaamste teughstanders van die geseyde predikant nie altemit nie een afgesproke werk nie. (1864, Jantje Eenvoudig: quoted in Wafer 1994a: 107)

Hulle gee vir jou absoluut geen handleiding nie op universiteit nie. (Feinauer 1994: 95)

niemand sal nie daarvan kry nie (Rademeyer 1938: 74)

In sum, the Afrikaans negation is believed to have developed 'on the back' of available dialectal patterns of (sentence internal) multiple negation, but is in itself an innovation (Ponelis 1993: 478).

A rather different hypothesis was put forward by Nienaber (1965/1994b) and Den Besten (1985/1986), who argued that the construction is the result of substratum influence from Khoe. In other words, Khoe speakers of Afrikaans developed this construction in analogy to their native language which has sentence-final negators, and the construction was later borrowed into the colonists' vernacular. Nama, for example, has three negative operators, two of which occur clause finally: tama (clause final, but may be moved leftwards) and tite (clause final). The

166), and Kotzé has described it for KVA (1983: 164). Leftward movement of negative markers has been reported for many languages (Horn 1989: 452), including pidgin/creoles (Van Rossem & Van der Voort 1993: 112).
prohibitive negative *taa*, however, usually occupies the first position (for details see Ponelis 1993: 470-471). Learning Dutch the Khoe shifted the locus of negation to the left, but maintained their substrate pattern of a sentence final negative element which was then reanalysed as a scope marker (Den Besten 1986: 221). Two main arguments have been brought forward against substratum influence from Khoe: (i) Nama negation is not strictly sentence final (i.e. the negator *tama* is often moved to the left, and the prohibitive negator *taa* is never sentence final), and (ii) Nie-2 is today facultative in the strongly Khoe-influenced varieties of Northwestern Afrikaans (ORA) but obligatory in other varieties (Ponelis 1993: 471-472).

The Middle Dutch negation with *en* and *niet* occurs in the CDV sources until about 1750, but is not believed to have been common in the spoken language (Ponelis 1993: 464). Examples of a sentence-final second negator surface in the documents from the early 18th century, but are scarce (Raidt 1983: 189, Ponelis 1993: 474-475). Furthermore, in the early CDV examples of a sentence final negator, Nie-1 is usually placed near the end of the clause.

\[(21) \quad \text{want ik heb geen vriende niet (Kasper Meester 1716, CJ 108/567)} \]
\[\text{de man segt immers geen quaad niet (Elise Botha 1762, CJ 1085/335)} \]

An interesting piece of evidence comes from the early 19th century, when C.J. Foecee (Fouché) transcribed two verbatim utterances by his Khoe farmhand Stoffel in a report to the magistrate (KT 206, 1810). A third negative within the main report conforms to norms of continental Dutch.

\[(22) \quad \text{wat hij niet weet niet} \]
\[\text{hij [...] zeg dat ik hem van avoon niet moet los maak niet} \]

Ponelis (1993: 475) who is careful not to overinterpret this marginal evidence, suggests that 'Foecee might be stigmatising Stoffel's speech by putting a vernacular stereotype in Stoffel's mouth', but rejects the possibility that this evidence could point to a real sociolinguistic

---

18 It has also been suggested that Nie-2 might be the result of interference from a dialect of Creole Portuguese, spoken on the west coast of Africa (Valkhoff 1966: 100-101). Wafer (1988), however, comes to the conclusion that although the two constructions show important similarities, there is no one-to-one correspondence between their negation systems, and the Afrikaans negation must be seen as structurally different. Ponelis (1993: 473) has furthermore pointed out that only 0.8% of the slaves came from the west African coast, and Asio-Portuguese dialects, which were far better represented at the Cape, lacked the clause final negator.

19 The negation particle in Middle Dutch was *en* or *ne* and usually preceded the verb. Increasingly post-verbal *niet* was used in addition to the preverbal negation particle. With the stabilisation of obligatory V-2 order in main clauses from the 16th century, the preverbal negator (which violated the V-2 structure) slowly disappeared, and the full weight of the negation was carried by post-verbal *niet* (Burridge 1992: 178-220).
difference, since 'there is clear evidence soon after 1810 of ample use of nie-2 in the speech of
whites'. However, this evidence falls almost entirely into the category 'dialect literature', which
is (as argued in § 3.1) useful in identifying the absence or presence of features, but generally fails
to capture the structure of linguistic variability. In 1830 the traveller Teenstra remarked with
regard to the negation in CDV 'deze herhaling van niet is hier zeer algemeen in gebruik', and
gives several examples of sentence final Nie-2 in his samenspraak (§ 3.1).20 Most of the texts
belonging to the vernacular writing tradition show examples of sentence-final Nie-2; however,
the frequency with which the vernacular writers use this feature varies greatly (see the

Based on Teenstra's evidence Nie-2 is believed to have been categorical in the spoken
language from the early 19th century (Raidt 1983: 190, 1991: 223 Ponelis 1993: 475), and the
much stretched argument of the conservative influence of the Dutch scribal practice is invoked
once again to explain both absence and variability in the historical record. However, while Nie-2
is an easy candidate for conscious avoidance, its excessive use can also be a stylistic choice to
give a text a 'vernacular flavour'. Considering that dialect writers tried to recreate the patterns
of the spoken vernacular in their writings, frequent use of Nie-2 (as well as other stigmatised
forms) can be expected, and the absence of a feature which was supposedly categorical in the
spoken vernacular demands an explanation. Lack of generalisation of Nie-2 as an obligatory
scope marker in these texts might actually point to a transition period in the 19th century, i.e.
Nie-2 was part of CDV but not yet generalised, and perhaps the sociolinguistic implications of
Foecees's evidence should be entirely discounted as an artefact. Note, that Nie-2 is entirely absent
from Johanna Duminy's acrolectal diary (Franken 1953) as well as from Trigardt's more
mesolectal diary (Smuts 1968). The lack or scarcity of documentary evidence can possibly be
explained to some extent by situational factors, and it was argued recently by Cheshire (1998
[1995]: 134) that negation is prominent mainly in conversational interaction:

I want to stress that negation generally [...] has an interactional role in ensuring the coherence of
the emerging discourse. In other words, it can link the current turn to the previous one, by negating
a presupposition that has just been expressed, whilst simultaneously ensuring that the interlocutors
have a shared orientation to the topic that they are pursuing, so that the subsequent turn is
felicitous. This cohesive function is perhaps one reason why negation occurs more frequently in
spoken than in written discourse: Tottie (1991) found twice as much negation in her sample of
spoken English as in the sample of written English.

Not only is negation generally more frequent in the spoken language, but one can also

---
20 Interestingly, the slave in Teenstra's dialogue uses not only Nie-2 but also Nie-1 in post-subject
position: hij niet spreek. Post-subject negation occurs also in Malay Afrikaans texts such as Abu Bakr's
Bayānudi (see Ponelis 1981: 76).
assume that certain negation strategies (especially those involving emphasis) will be typical of spoken rather than written discourse. An example of this comes from Boniface's vernacular dialogues with Hendrik Kok (see § 3.1), where at the end of the second dialogue the interviewer accuses Kok of alcohol consumption at the house of the philanthropist Phillips. Kok denies this emphatically with a multiple negative (Nee, niks geen drank, nie, niks) which is unlikely to have survived in conventional written prose (that Nie-2 relates to discourse-dependent structures in the basilectal varieties was also suggested by Roberge 1998: 35)

Discussion of the origin and development of Nie-2 is further complicated by the patterns of contemporary non-standard usage where Nie-2 is often omitted, or additional Nie-3 appears as a result of the speakers uncertainty of where Nie-2 should be placed (see Tiflin 1984: 73ff., Van Rensburg 1989: 147, Links 1989: 71, Feinauer 1994). In a recent re-assessment of the history of Nie-2, Raadt (1994 [1992]: 305) remarked cautiously that in the light of such contemporary patterns historical linguists might have been overconfident regarding their rather early dating of Nie-2.

Dit wil amper voorkom van die negasie patroon van die sogenaamde 'dubbele ontkennin' tog nie so vas en sonder uitsondering was nie. Veral in die lig van hedendaagse gebruikspatronne het ek iets wat versigtiger geword in die interpretasie van ouer gewens. Hoewel die ontkennin met eind-nie as tweede negasiepartikel voor 1875 al sekerlik die normale negasiepatroon was, kon ontkenningsvariante sonder die eind-nie ook gebruik word.

Leaving open for the time being the question which negation pattern was the 'normal' one (in the sense of quantitatively dominant), pushing the stabilisation of Nie-2 as an obligatory scope marker into the second half of the 19th century implies the possibility that obligatory Nie-2 might be a standardisation feature, an interpretation which is supported by the existence of variable multiple negation patterns as well as lack of Nie-2 in non-standard dialects of Afrikaans (see Ferguson 1995: 179 for a similar argument explaining the categorical use of third person singular -s in standard English as a standardisation feature).

Support for the interpretation that historical Nie-2 originally functioned as an emphasis-marker comes from the writing convention adopted by Boniface, who regularly separated Nie-2 from the rest of the clause by a comma.

(23) Droogkeel: Ach! moet toch nie bakkelyen, nie. (De Nieuwe Ridderorde of De Temperatisten, 1832)

Nienaber (1994b: 109-110) has interpreted the comma as a stylistic device, aimed at emphasising the second negator by drawing the reader's attention to it. Yet, keeping in mind that the primary function of a comma is to indicate a pause between constituents or to divide items
in a list, that is, to disassociate, the comma can be seen as separating Nie-2 from the sentence core, possibly assigning it to the role of an emphasising particle (similar to '..., eh?' in colloquial speech). However, if Nie-2 ever had an emphasising function (this possibility is also hinted at by Burridge 1992: 201, fn.9), this was lost when it became grammaticalised into a scope marker, as today Nie-2 is a as rule unstressed (Wafer 1978: 64, 203).

The traditional (descriptive) model for the change of negation structures is based on Jespersen's (1917) argument that changes in the structure of negation typically show a cyclic pattern:

The history of negative expressions [...] makes us witness the following curious fluctuation. The original negative is first weakened, then found insufficient and therefore strengthened, generally through some additional word, and this in turn may be felt as the negative proper and may in due course of time be subject to the same development as the original word. (1917: 221)

Thus, for Dutch (from Old Dutch to Modern Dutch) we get the following development (see Horn 1989: 454-455, Burridge 1992: 200ff., Hoeksema 1995):

preverbal negation  →  brace negation  →  postverbal negation
ne/en/n  →  ne/en/n... niet  →  niet

It is possible that the Afrikaans development relates to this general pattern through the introduction of a second negative which was originally emphatic. That is, as Nie-1 moved leftwards, its ability to mark the scope of negation was weakened as a result of its variable position (see Ponelis 1993: 478); it was then reinforced by rightward Nie-2, which was later reanalysed as a scope marker. The agents of this re-analysis were certainly not only the colonists but also the Khoe and slaves, who, one might speculate, had been exposed to a fair amount of final negation through imperative sentences of the type Slaap nie, Droom nie (as well as pidgin-type constructions used by substrate speakers such as slage nie, literally 'blows not', i.e. 'Don't hit me'; see Roberge 1998: 35).21

Examples of Nie-2 are relatively rare in the corpus and only 21% of negative clauses show the Afrikaans brace negation (363 out of 1746 negative clauses). The distribution across cases (figure 7.3) is extremely skewed with only about a third of individuals employing the feature at all. Use of Nie-2 was not yet categorical for any informant.

---

21 There is no indication that an any stage Nie-2 was interpreted as the negative 'proper' which alone could carry the weight of the negation. Use of only a sentence final negator has been reported for Berbice Dutch, a Dutch-lexicon creole. In Berbice Dutch the standard negation contains only the unstressed and sentence final negator ka(ne) which is related to the Eastern Ijo negator ka (also sentence final) and the Dutch negator nee. Berbice Dutch has also maintained a type of brace negation in which a preverbal negator (noko), a negative adverb or a negative verbs can occur clause internally in addition to sentence-final ka (Kouwenberg 1993).
Figure 7.3 Histogram for 85 cases, summarising the use of Nie-2 across cases (based on percentages for individuals).

Those individuals who employ Nie-2, use it in agreement with the grammatical rules of Afrikaans. Amongst those who use Nie-2 relatively frequently there is, however, a tendency to omit the second negative in cases where several constituents would intervene between the two negators, especially if the position of Nie-2 would be at the end of a complex subordinate clause. It might be that such constructions are partially constrained by the V-final rule in subordinate clauses, which would be upset by final Nie-2. Examples of (Germanic) negative spread also occur in the data.

(24) hy heef nooit geen vout gemaak (M.C. Hillebrand, 13/10/1903) de schoolkamer zoo klein is dat hy onmogelyk niet voor 23 kinderen kan school houden (Elizabeth Malan, 9/5/1893)

Examples of Nie-3 although not very frequent, nevertheless occur in the data:

(25) ik glo niet dat hulle iets zal doen nie voor dat de Vrystaat zyn volksraad zit nie (D.C. DeWet, 24/11/1888) was nie lang genoeg nie om die Afrikaanse beweging van alle kante te bespreek nie, en makaar syn standpunt te verstaan nie (F. S. Malan, circa 1895)

In sum, while Nie-2 is likely to be old since early examples conform to the structural principles which characterise Afrikaans negation, generalisation of the structure as a negation 'rule' might have been relatively late. Like the demonstratives hierdie and daardie, Nie-2 might be described as a habitualised pattern with relatively low frequency, possibly socially marked
(since it does not occur in acrolectal and upper-mesolectal texts), and occurring most typically in face-to-face interactions.

### 7.3.2 The Infinitive with om...te

In Dutch three types of infinitive clauses can be distinguished: (i) bare infinitives, (ii) constructions with the infinitival marker te, and (iii) constructions including the complementizer om as well as the marker te. Infinitives with om...te are obligatory only in cases where the infinitive phrase either (i) describes the head noun further (predicative nominative), or (ii) clarifies the aim or intent of an action (that is, constitutes an adverbial of purpose; Geerts et al. 1984: 716-717, 789-90, 824, Robbers 1997: 47-51). Bare infinitives can occur in subject position, but also follow modal verbs and some copula verbs (as in Zij hoorde mij komen). In all other cases te-infinitives are acceptable in Dutch.

\[(26)\]

(i) Bare infinitives

Slapen is gezond.

(ii) te-infinitives

Laura was blij met haar moeder te reizen.

(iii) om...te - infinitives

Het is een meisje om te zoemen.

Ik vraag het niet uit nieuwsgierigheid maar om je te helpen.

In spoken Dutch om...te infinitives frequently replace te infinitives and the distribution of om...te and te infinitives is said to vary socially and stylistically, with om.te being widespread in informal registers (De Voors 1953: 163-164, Donaldson 1981: 158, Geerts et al. 1984: 790, Ponenis 1993: 292).\(^{22}\) Afrikaans has generalised the om...te construction and infinitives with te are largely obsolete in spoken Afrikaans. Bare infinitives in subject position are also frequently paraphrased as om...te constructions.

\[(27)\]

Slaap is gesond. - Om te slaap is gesond.

Laura was bly om met haar moeder te reis.

Te-infinitives have survived in idiomatic expressions (for examples see Raidt 1983: 177f.), and the modals behoor(t), hoef and durf as well as the copula 'to be' usually combine with te (Ponenis 1979: 432, Donaldson 1993: 274-278, Robbers 1997: 89). In more formal styles the verbs skyn, blyk, dien, meen and wens are often followed by te infinitives (Ponenis 1979: 248). Infinitives with te occur also in relative structures such as Hy het 'n verhaal te vertel (see Ponenis 1993: 293). However, it appears that the use of te-complements is decreasing, and they are

---

\(^{22}\) Om...te infinitives replace te infinitives also frequently in Frisian (Hoekstra & Tiersma 1994: 525) and (spoken, informal) German (um...zu infinitives).
generally perceived as archaic and formal (Robbers 1997: 214). Bare infinitives follow in Afrikaans the modals *kan, mag, moet, sal* and *wil*, and are also common with copula (linking) verbs, as in *Dit begin nou reën* (Donaldson 1993: 273-274, Robbers 1997: 56ff.).

The syntactic structure of the *om...te* infinitive in both Dutch and Afrikaans can be summarised as:

\[ om + (X) + te + (\text{infinitive}). \]

The position of \((X)\) can remain empty or can be filled with an NP, PP or adverb. In cases where the verb has a separable prefix, the particle *te* is positioned between prefix and stem (*opstaan < om op te staan*). Non-standard dialects of Afrikaans show considerable variation in the structure of the infinitival complement. This has been described in detail for Griqua Afrikaans (see Webb 1993: 168-69):

\[
\begin{align*}
(28) \quad om \ te + (X) + (V) & \quad om \ te \ werk \ soek (< om \ werk \ te \ soek) \\
om \ te + (X) + te + (V) & \quad om \ te \ hier \ te \ kom (< om \ hier \ te \ kom) \\
om + (X) + (V) & \quad om \ die \ waarheid \ sê (< om \ die \ waarheid \ te \ sê) \\
om + (X) + te + (\text{particle})+ (V) & \quad om \ die \ bees \ te \ vasmaak (< om \ die \ bees \ vas \ te \ maak)
\end{align*}
\]

Roberge (1994b: 87-89) has suggested that such non-standard infinitives originate from basilectal infinitive constructions which only used the complementizer *om*.\textsuperscript{23} Mesolectal varieties, which were under the influence of the Dutch acrolect, added the particle *te* in different syntactic positions (as exemplified in the Griqua data).

If I am correct [...] convergence between acrolectal Cape Dutch and the Cape Dutch Creole resulted in two hybridizations. Basilectal complementation with *om* was made formally more similar to its acrolectal counterpart through the introduction of *te* first in tandem with *om* and then secondarily into the position directly before the infinitive itself. (Roberge 1994b: 88-89)

The generalisation of *om...te* infinitives in acrolectal CDV, on the other hand, can be linked to inherited Netherlandic patterns (see above for the social/stylistic variation patterns in spoken Dutch, also Raidt 1983: 179).

In order to investigate the spread of the *om...te* construction into the domain of *te* infinitives in the corpus, only those infinitive constructions in which use of *om...te* was not obligatory according to the Dutch system were tabulated and the percentage of *om...te* infinitives was established (bare infinitives were excluded). 1200 infinitive clauses were tabulated, 618 of

\textsuperscript{23} Note, however, that infinitives with the structure *om + (X) + (infinitive)* are also attested for Early Dutch and the southern Dutch dialects (Ponelis 1993: 292, LeRoux 1923: 104).
which made use of the complementizer *om* (52%).

Figure 7.4 Histogram for 74 cases, summarising the generalisation of the *om...te* infinitive across cases (based on percentages for individuals).

Compared to distribution patterns for the other variables, the lack of polarisation is remarkable, indicating that the majority of individuals made moderate to frequent use of *om...te* infinitives in positions where the (prescriptive) norms of standard Dutch would have favoured only *te*. Thus, on the whole the construction is common in the data without having been entirely generalised, and *te* infinitives still occur. *Te*-infinitives are categorical in the corpus after the modals *hoeven*/*hooft* and *behooren*/*behoort*) (12 and 5 instances respectively), which is also the case in Afrikaans and Dutch (Donaldson 1981: 152, 154, Geerts *et al.* 1984: 535-536).

(29)  *en ek hooft nie te se dat dit heerlik is* (N.J. De Wet, 12/5/1895)  
*Krause behoort 'n First Class gekregen te hebben* (F.S. Malan, 1/6/1893)

Interesting is the use of the verb *begin* + infinitive, which is followed in the corpus generally by *te* + (V) or a bare infinitive, and only one example of *begin* + *om...te* was found. In Afrikaans *begin* is usually followed by a bare infinitive, but can also be followed by *om...te* (Donaldson 1993: 274, 277). In Dutch on the other hand *beginnen*, although functioning as an aspectual auxiliary, is generally followed by *te* (e.g. *Jan begint zijn viool te stemmen*). Use of a bare infinitive is possible in some Dutch dialects if the infinitive directly follows *beginnen* (*Ik moet beginnen werken*, *Ze beginnen vandaag werken*; see Van den Toorn 1982: 193, Geerts *et al.* 1984: 585).
(30) \( \text{begin} + \text{te} + (V) \)
Zaterdag het dit weer \textit{begint te} reent (Johanna Brümmer, 29/3/1895)
de vogels \textit{begin al te} eet er aan (H.S. Vermaak, 19/2/1906)
\textit{begon} de mensen \textit{te} huilen (W.C. Cronje, 19/7/1905)

(31) \( \text{begin} + (V) \)
de tuin \textit{begint} prachtig \textit{worden} (B.J. Brümmer, 19/9/1907)
ek het dadelijk \textit{begin} geld bij mekaar \textit{maak} om 'n klavier vir ons skole te kry
(Lelom Malan, 31/10/1917)
ek \textit{begin} al so stadig die motor \textit{nasien} (A. Hauptfleisch, 7/9/1921)

(32) \( \text{begin} + \text{om te} + (V) \)
wil ik by de Einde van deze maand \textit{begin om te snij} (E. Smith, c.1900)

The complementizer \textit{om} occurs also as a causal conjunction in the corpus, replacing the
'proper' form \textit{omdat} ('because'). Use of \textit{om} in this function has been described for earlier CDV
texts (Roberge 1994b: 87), and was branded by Malherbe (1917: 104) as colloquial, and strongly
discouraged.

(33) Pa moet voor hem een wijzig geld \textit{stuur om} hij niets heeft (Petronella van
Huyssteen, circa 1906)

Non-standard infinitives of the type illustrated in (28) do not occur in the corpus, yet were
found by chance in a letter (late 19th century) which was written by an English speaker who had
just begun to learn Dutch (SAL-MSC1-34-9). The letter was not included in the main corpus.

(34) \( \text{om} + (X) + \text{te} + \text{particle} + (V) \)
ik heeft ooit probeerd \textit{om te oplos}
\textit{om te} + (X) + (V)
\textit{om te} zien zoovele arme kinderen

The fact that such non-standard infinitives are absent from the corpus, but occur with
great frequency in a short letter written by an L2 speaker, suggests the possibility that these
constructions are typical 'mistakes' of L2 speakers; however, more evidence from contemporary
L2 varieties of Dutch is needed to assess this interpretation.

7.3.3 Objective \textit{Vir}

The use of \textit{vir} (Dt. voor) as an object marker is typical of (especially) spoken Afrikaans,
where it can not only replace prepositions such as \textit{aan} or \textit{na} (Dt. \textit{naar}), but also occurs in
constructions where the equivalent sentence in Dutch does not allow the use of a preposition. The
use of \textit{vir} is obligatory with prepositional objects and in bi-transitive constructions where the
direct object is followed by the indirect object (35(i)). In such cases vir can be replaced by the 'historically correct' Dutch preposition in formal registers. Vir is facultative in bi-transitive constructions with the indirect object preceding the direct object, and before direct objects in general (35 (ii); see Carstens 1991: 59-60, Donaldson 1993: 342-343, Ponelis 1993: 265-266, Raidt 1994 [1969]).

(35)
(i) Obligatory objective vir
   Ek luister vir (na) Oom Jan.
   Hy het die boek vir (aan) sy broer gegee.

(ii) Facultative use of vir
   Hy het (vir) sy broer die boek gegee.
   Ek sien (vir) Nelsie.

   (Dt. Ik luister naar Oom Jan)
   (Dt. Hij heeft het boek aan zijn broer gegeven)
   (Dt. Hij heeft zijn broeder het boek gegeven)
   (Dt. Ik zie Nelsie).

Pronouns in sentence initial position show a strong preference for vir (as in Vir hom vra ons niks). Use of vir is optional with proper nouns and pronouns that are scrambled (i.e. that appear to the left of adverbs, as in Hulle het (vir) Sarie gister gesien), but obligatory when they are not scrambled (e.g. Hulle het gister vir Sarie gesien. *Hulle het gister Sarie gesien; Ponelis 1993: 268). The use of facultative vir in contemporary Afrikaans is governed by the nature of the noun (hierarchy: proper nouns (almost always) > pronouns (frequently) > substantives (seldom)), animacy (inanimate nouns do not usually take vir)\(^24\), and emotivity and degree of formality (emotional involvement favours use of vir; Ponelis 1979: 202-203, 1993: 266ff.).\(^25\)

There are some marginal (Dutch and German) dialectal structures which suggest that objective vir is not 'entirely without a Germanic antecedent' (Ponelis 1993: 271, see also Raidt 1994 [1976], LeRoux 1923: 140ff.). Examples of Dutch voor exist in bi-transitive constructions (indirect object following direct object) as well as with prepositional objects (bi-transitive: Mijn Oupa heeft een trui voor me gebreid, prepositional: ik zal voor hem betalen), and in Early Modern Dutch the preposition voor occurred in many contexts where modern Dutch requires different prepositions (for a detailed discussion and examples see Raidt 1994 [1976]). However, the evidence is scarce for those environments where Afrikaans allows facultative vir, especially preceding direct objects, and on the whole examples of analogous dialectal constructions are

\(^{24}\) The use of vir + inanimate object pronouns has however been described for KVA (Kotzé 1983: 256f.).

\(^{25}\) There is some debate surrounding the question as to whether facultative vir should be interpreted as a preposition, or whether it is more adequately described as a 'case prefix' or object marker (Den Besten 1981: 156-158, Markey 1982: 182; see Ponelis 1993: 269-271 for an argument in favour of the preposition status of vir).
few. Since the construction is also not attested for colonial Dutch, Afrikaans historical linguists have interpreted objective *vir* as an ET innovation triggered not so much by the Germanic input but by substrate influence from Creole Portuguese (Raidt 1994 [1976], Ponelis 1993: 272ff.). The similarity of the Low Portuguese construction *per* + personal object to the Afrikaans construction with *vir* was first pointed out by Hugo Schuchardt (1891). In Creole Portuguese both direct and indirect objects (including personal pronouns, proper names and substantiv phraces) can be preceded by the preposition *per*.

(36)  

nos atja ponta per eli  
WE CAN SHOOT FOR HIM (Raidt 1994 [1976]: 133)

The substrate hypothesis is supported by the fact that early attestations of objective *vir* occur predominantly in the speech of slaves, most of whom were conversant in Creole Portuguese (for examples see Raidt 1994 [1976], see also Ponelis 1993: 273f.). It is generally agreed that objective *vir* was a well entrenched feature of CDV by the beginning of the 19th century, and Changuion described in the mid-19th century the widespread use of the construction in *De Nederduitsche Taal in Zuid-Afrika hersteld*:


The last example given by Changuion suggests the possibility that objective *vir* was a social marker, as the image of the child-beating rough threatened with violence in return is certainly suggestive if not stereotypical, and according to Raidt's interpretation (1994 [1969]: 114, 1994 [1976]: 140) the use of objective *vir* was socially stratified. This interpretation is also supported by the (often emotional) discussions surrounding the translation of the Bible into Afrikaans, where the use of objective *vir* was a much contested issue (see Raidt 1994 [1976]: 141ff.). Frequent use of *vir* was typical for many vernacular texts in the late 19th century, although important differences existed between individual writers. While, for example, S.J. du

---

26 According to Schuchardt (cited in LeRoux 1923: 141) the construction *er sagte für mich* is possible in the dialect of Thuringia. In the Cologne urban vernacular, which shows a tendency to replace the dative with accusative, similar constructions appear to be possible. Thus, when replacing the dative in directional constructions (such as in *Gib das mir* with the accusative (*mir → mich*), the directionality can be emphasised by inserting the preposition *für*. *Gib das für mich*. Note that this construction is certainly not part of the old traditional Cologne dialect (now codified in grammars and dictionaries), but the spoken colloquial, especially amongst the younger generation. My own memory of such constructions is admittedly a bit hazy (having left Cologne at the age of nineteen) but my father, who has been a highschool teacher in Cologne for the last 35 years, agreed that such constructions are possible and not uncommon in the regional colloquial.
Toit used *vir* almost excessively in his writings, others made little use of the construction, often limiting its use to the reporting of direct speech, especially when representing people of colour (see Raidt *ibid*).

Altogether, 237 instances of objective *vir* (with the variants *voor* and *ver*) were found in the corpus, and 65 individuals (43% of the sample) made use of the feature. Since it was difficult to limit the pool of possible variants in a satisfactory way, objective *vir* was quantified by calculating its occurrence per 1000 words (see § 4.2). Most of the individuals used *vir* sparingly; frequent use was, however, a characteristic of the 'Afrikaans' texts of many of the bi-dialectal writers (see § 4.1.2). The tendency to use objective *vir* more frequently before proper nouns than before other nominal categories is noticeable in the corpus (example 37).

(37) groet die kinderlies en ver Frans (B.J. Brümmer, 28/11/1907)
beste groete aan almal en ook vir Katte van Winnie (Wynanda Hoogenhout, 13/3/1911)

Frequent (27 instances) was the construction *jammer vir iemand wees* which was reported by Changuion in his *Kaapsch Taaleigen* (1844: xiv), and remarked upon by Scholtz (1965: 93) in his analysis of the Hendrik Kok dialogues as *eienardig*. Widespread was also the use of the copula *lyk* plus *vir* followed by what Ponelis has termed an 'experciener NP' (as in *dit lyk vir my*; see Ponelis 1993: 275-276, Raidt 1994 [1969]: 95). The quantitative analysis presented in table 7.12 and table 7.13 indicates that use of *vir* was at least partially dependent on the lexical verb involved, a tendency also known in contemporary Afrikaans, where certain transitive verbs commonly combine with *vir* (see Raidt 1994 [1969], Carstens 1991: 60). Surprising is the low frequency of *vir* for *geven/gee*, *vragen/vra* and *zeggen/zê* which today typically combine with *vir*. However, although lexically restricted, objective *vir* is well established in the corpus.

<table>
<thead>
<tr>
<th>verb</th>
<th>objective <em>vir</em> (N)</th>
<th>total</th>
<th>% of objective <em>vir</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>sturen</td>
<td>34</td>
<td>82</td>
<td>42%</td>
</tr>
<tr>
<td>schrijven</td>
<td>20</td>
<td>124</td>
<td>16%</td>
</tr>
<tr>
<td>zeggen</td>
<td>12</td>
<td>113</td>
<td>11%</td>
</tr>
<tr>
<td>vragen</td>
<td>7</td>
<td>67</td>
<td>11%</td>
</tr>
<tr>
<td>vertellen</td>
<td>3</td>
<td>34</td>
<td>9%</td>
</tr>
<tr>
<td>geven</td>
<td>8</td>
<td>133</td>
<td>5%</td>
</tr>
</tbody>
</table>

Table 7.11 Use of objective *vir* with different lexical verbs preceeding direct and indirect objects (C-1880-1922). $\chi^2 = 52.57771$, df=11, p< 0.001.

27 Note, however, that case assignment is often rather fuzzy in conjunction phrases. In dialectal English this phenomenon is well attested, i.e. *Mary and I/me played in the garden* (for further details see Luijks 1998).
7.3.4 The Periphrastic Possessive with 'Se'

Apart from the use of attributive and predicative possessive pronouns to indicate possessive relationships (see §7.1.2), three other constructions are available in standard Dutch to express possessive relations: the sigmatic genitive (-s), the van construction and a periphrastic construction with the pronoun zijn, z’n/haar, d’r (which varies according to gender and number of the possessor; Geerts et al. 1984: 53). 28

(38) sigmatic genitive
     van construction
     z’n/d’r construction
moeders huis
het boek van mijn vriendin
mijn vriendin d’r/haar boek
mijn vriend z’n/zijn boek

While the sigmatic genitive had already disappeared in Afrikaans by the 18th century (Scholtz 1963: 108), the van construction is still used in modern Afrikaans but not very common (Donaldson 1993: 99). A periphrastic possessive similar in surface structure to the Dutch d’r/ z’n construction is, however, highly productive in Afrikaans. The Afrikaans construction uses an invariant form se which can follow not only animate nouns (as in Dutch) but also inanimate nouns, personal pronouns as well as other noun phrases including expressions of measurement and time (see Donaldson 1993: 98-100, Ponelis 1993: 231-239, Roberge 1996: 123-125; for a generative account see Oosthuizen & Wafer 1994). Periphrastic possessives with se are perfectly acceptable in spoken and written Afrikaans (Donaldson 1993: 98, Ponelis 1993: 241), while the z’n/ d’r construction in Dutch is colloquial, although no longer limited to the spoken language (Geerts et al. 1984: 208-209).

---

28 In spoken Dutch the weak form of the pronoun is typically used; however, Roberge (1996: 148, n.) refers to a conference paper by Muriel Norde in which she argues that strong forms have made inroads into spoken Dutch (Norde, M. (1995): 'Grammaticalisation versus reanalysis: The case of possessive constructions in Germanic' Paper presented at the Twelfth International Conference on Historical Linguistics, University of Manchester, 13-18 August 1995).
The use of personal pronouns as possessors is not part of the system of standard Afrikaans, but widespread in non-standard varieties of western Cape Afrikaans (Ponelis 1993: 231) as well as ORA (Rademaker 1938: 66-67, see also Malherbe 1917: 83). Afrikaans, furthermore, allows for the insertion of postmodifying clauses between possessor and possessive pronoun (as in die man met die mooi huis se kind) and the construction is (potentially infinitely) recursive (as in die man se dogter se vriendin se boek...).

The Afrikaans periphrastic possessive with se is not only more productive than the Dutch z'n/ d'r construction, but differs structurally from the Dutch pattern as the invariant form se is not a possessive pronoun but a free possessive particle which cliticises to the preceding NP. In other words, the possessor NP + se forms a determiner which specifies the possessed noun (the head of the phrase; Le Roux 1923: 89, Den Besten 1978: 31-38, see also Donaldson 1993: 99, fn.54).

Although Afrikaans se has been interpreted in the literature initially as a spontaneous reanalysis of the Dutch pronominal construction which was triggered by the loss of gender agreement in the nominal system (Scholtz 1963: 107-108), it is today accepted that the development of the structure owes much to language contact and convergence, and strikingly similar constructions have been reported for Creole Portuguese, Malay and Nama (Le Roux 1923: 97-98, Valkhoff 1966: 17-18, 227-229, Den Besten 1978: 31, Ponelis 1993: 225). All three languages employ a possessive which is periphrastic in structure, contains a free and invariable possessive particle, shows the same order of elements, and allows noun phrases as well as personal pronouns as possessors. The Creole Portuguese possessive (sua, weakened to su) allows

---

25 Note that some dialects in Flanders also make use of an invariant possessor (Taeldeman 1994); however since Holland dialects dominated in the dialectal base of Afrikaans (§ 2.2.1) direct dialect influence is unlikely.
furthermore for the expression of temporal and other relations in ways similar to Afrikaans (Ponelis 1993: 246).

(40)
Creole Portuguese

me sua nomi
I POSS name

minya sua
I POSS

my

Malay

Dodi punya rumah
Dodi POSS house

saya punya rumah
I POSS house

my house

Nama

khoeb di haab\(^{30}\)
man POSS horse
the man\'s house

sada di haab
we POSS horse
our horse
(examples from Ponelis 1993: 244-45)

The lexical form se has been interpreted as a reduced allomorph of the Dutch pronoun zij(n). Ponelis (1993: 235), for example, explains se as a result of atomic reduction and nasal deletion: *sein* → *sen* → *se*, and assumes that (despite its scarcity in the corpus of 18th century CDV where *sen* and *syn* are the common forms), se must have been part of the spoken language in the 18th century (also Scholtz 1963: 109). Roberge (1996), however, has argued that to explain se as an allomorph of early modern Dutch *zijn* could be 'barking up the wrong tree' because se is so very rare in the sources before the 1850s, and its origin might, therefore, lie outside the acrolectal Dutch system. In other words, se could have been derived from the reduced Creole Portuguese possessive formative *su* (*su(w)a* → *su* → *se*). According to Roberge the use of the invariant particle se was initially limited to basilectal varieties, but later converged with the Dutch based periphrastic possessive of the European settlers.

The use of the Dutch gendered emphatic forms (*syn* and *haar*, as well as *hulle* for the plural) was still dominant in 18th century CDV, and persisted until the early 20th century (for examples and discussion see Ponelis 1993: 232-37, Roberge 1996: 135-39; see also Elffers 1903: 12, who still reports *s\'n* as the singular form (masc.and fem.), and *hulle* as the plural form). Occurrences of the masculine possessive with feminine or plural antecedents, on the other hand, were rare during the first half of the 19th century, but became increasingly common after 1850

---

\(^{30}\) The genitive particle *di* is, however, omissible in Nama, which makes direct substrate influence less likely. Roberge (1996: 134, 141) has argued that the possibility of a particleless possessive structure in Nama, might have led to the juxtaposition of possessor and possessed noun in some basilectal varieties. He cites the following construction from a letter written by 'a Griqua' to the *South African Commercial Advisor* (1830): *in ander menschen land* (in other people's land). Similar constructions were found by Luijks (1998) in the letters of the Nama leader Jan Jonker Afrikaaner: *God woorte* (God's words), *my Vader blads* (my fathers place).
(for a detailed discussion of the diachronic development see Roberge 1996: 135ff.). However, grammatical descriptions of Afrikaans as late as the early 20th century still report the Dutch forms sy(n), haar and hulle/hul as being widespread in the spoken language (Malherbe 1917: 83, Le Roux 1923: 88-9, Bouman & Pienaar 1925: 76, Scholtz 1963: 108).

52 of the sampled individuals made use of the periphrastic possessive in their writings (38% of the sample), and a total of 148 periphrastic constructions were tabulated. Of these 55 had either a female or a plural possessor, which would have required agreement in the possessive formative according to the Dutch system; however, just over half of these (N=30, i.e. 55%) combined with the correct plural or feminine possessive pronouns. The majority of periphrastic possessives in the data have [+animate] nouns in the position of the possessor, while nonanimate nouns occur but are rare. Only three examples were found in which a pronoun fills the slot of the possessor.

(41) **Possessor: NP [ + female]**

twee van Eliza zyn broeders (Johan Brümmer, 12/7/1900)
Hester zyn schoolgeld (H.H. Pienaar, 17/5/1905)

(42) **Possessor: NP [ - animate]**
die zoo z'n postkaartjes (Jacques Malan, circa 1910)
Queenstown se kerk (Hlela, 29/9/1906)

(43) **Possessor: NP [ + plural]**
zyn kenders zyn gesonheid (B.J. Brümmer, 14/10/1904)
die boontjies se kost (Bettie Smit, 23/11/1917)

(44) **Possessor: NP [ + pronoun]**
met hun allen zyn boerderij (B.J. Brümmer, 26/4/1902)
mekaar zyn standpunt (F.S. Malan c.1895)
mekaar se gesondheid (Grosskopf, 26/3/1907)

There are some examples in the data which suggest that there might have been a stylistic constraint regarding variation between the invariant particle and the use of gendered and plural pronouns. Hester van der Bijl, for example, uses mostly the particle se in her rather 'chatty' vernacular letters (*my ma se soetste kind, Dr Klaas se stoute, sy Pa se ouders*), yet when writing within the same corpus of letters about the recent death of Johanna Hoogenhout, she uses the correct Dutch pronominal form (*Johanna haar dood, also Mijinheer Roussouw zyn dood*). Rachie Malan uses the construction with se in a letter (8/5/1909) which otherwise conforms entirely to the norms of continental Dutch, and other 'Afrikaans' markers such as inflectional reduction, loss of gender, nominative ons, Nie-2 or objective vir are absent. The construction quoted in (45) occurs in the context of a short anecdotal description of the monolingual limitations of English speakers who cannot make themselves understood on a German ship.

(45) het wekt 'n mens se lachspieren op
In sum, although the data show a tendency towards loss of NP constraints, the development towards an invariant possessive formative is not yet far advanced, and many writers still show control of the Dutch system.

7.4 Comparing Distribution Patterns

It was argued in §5.1 that descriptive statistics which summarise the distribution of cases (= individuals) across data values, are valuable as they can provide leads for the formulation of hypotheses and thus further investigation. Comparing the histograms for the morpholexical variables with those for the morphosyntactic variables (pronoun variables and het) different distribution patterns are noticeable: for the morpholexical variables the distribution of cases is polarised with relatively few cases in the middle field, while in the histograms for the morphosyntactic variables (loss of gender and regularisation of the verbal inflection) polarisation interacts with considerable continuity across the middle field. The distribution is skewed for Nie-2 (with the majority of individuals not making any use of the Afrikaans brace negation, positively skewed). The direction of a skew in a distribution depends on the value of the mean and the median: a distribution in which the mean value is higher than the median value is said to be 'positively skewed'; a distribution in which the median value is higher than the mean value is 'negatively skewed'. A skewed distribution was also found for the loss of the preterite (with the majority of individuals showing scores of over 60%, negatively skewed). Lack of polarisation is noticeable in the distribution pattern for the om...te infinitive, which is characterised by a general and continuous distribution across the middle field. The strongly polarised structure of most graphs can be interpreted as indicating the existence of two relatively focused norms in the Cape speech community, i.e. the old target 'Dutch' and the new or emerging target 'Afrikaans'.

An important question is whether and how such distribution patterns across speakers in apparent time relate to the diffusion of a linguistic change in 'real' time. It could be argued, for example, that strongly polarised distributions without much variation in the middle field reflect the existence of two largely separate norms, while continuous distributions reflect processes of slow and gradual spread across social groups (or possibly stable variation). Skewed distributions in which cases lie disproportionately at one end of the range of values could be interpreted as indicating either very late (negatively skewed distributions) or early (positively skewed distributions) stages of a change.

To assume that the patterns of diffusion across speakers at time t allow us to generalise

---

31 Apparent time studies describe the spread of a linguistic variant at point t across different age groups, and are based on the assumption that the speech of younger speakers is advanced further than the speech of older speakers, whose language is assumed to resemble the vernacular of their youth. In other words, change in apparent time is believed to relate directly to change in real time.
about the temporal diffusion of the linguistic change under description also opens the question of how these histograms relate to S-curve graphs, which have been shown in many studies to adequately describe the temporal diffusion of innovations (linguistic and others; for a general overview see Mahajan & Peterson 1985). The S-curve model predicts slow spread in the initial stages of a change, followed by a period of (rapid) acceleration in which the change catches hold, and finally the trajectory of the diffusion slows down before it is completed. Chambers & Trudgill (1980: 178-180) have argued that the S-curve model can also be applied to the description of change in apparent time by plotting speakers (rather than dates) on the horizontal axis (the vertical axis remains unchanged, indicating the percentage of usage of a linguistic variant). It is then hypothesised that those speakers which show low percentages 'will move rapidly into the middle areas and into the high percentages in the course of time' (Chambers & Trudgill 1980: 180). Similarly Wardhaugh (1992: 212-213) suggested that an apparent time curve which closely approaches the 'ideal' S-curve can be interpreted as describing a change in progress which has reached its mid-point.

Figure 7.5 which plots the data for loss of gender across cases, is easily reconcilable with the S-curve model. About one-third of the individuals show loss of gender in no more than 20% of possible cases, one third shows loss of nominal gender in 21-80% of possible cases, and one third in 81-100% of possible cases. In an ideal S-curve the proportions should be roughly: 37.5% of cases ≤ 20%, 25% of cases 21-80%, 37.5% of cases 81-100% (Bailey 1973: 77). This indicates, according to Wardhaugh's interpretation, a change in progress; roughly at mid-point.

![Graph](image)

Figure 7.5 Loss of neuter gender across speakers (apparent time).

However, while the S-curve model has been shown to be of theoretical importance for
our understanding of diachronic change (allowing not only for description, but also prediction), it is by no means clear that the patterns shown to exist across time are also of importance for our understanding of diffusion in apparent time. This was emphasised by Devitt (1989: 48), who argued on the basis of her analysis of the diffusion of English linguistic norms in Scotland (1520-1659) that 'there is no necessary connection between S-curves in apparent time and S-curves in real time', and that although variables might not show an S-curve in apparent time, their historical diffusion nevertheless follows the logistic model. Devitt's research also showed that finding a logistic curve for two separate variables at time $t$ probably tells us little about further rates of change. In her study, for example, the apparent-time graphs for the spread of the negative particle and the indefinite article appear quite similar; their development in real time, however, proceeded at different rates: while the indefinite article anglicised only slowly, the negative anglicised at a very rapid rate.

In sum, when interpreting distribution patterns across speakers in apparent time, it is necessary to acknowledge that little is still known about how these patterns relate to the emergent macro-structure of the temporal S-curve. To postulate that diffusion patterns across time will be replicated in apparent time studies ignores the fact that the mathematical diffusion model describes by definition the temporal aspects of a diffusion (rates of adoption across time), but not (necessarily) the geographical or social diffusion of the innovation which might follow rather different patterns (Mahajan & Peterson 1985: 25, 41-43). However, although the relationship between apparent and real time is certainly 'more complex than a simple equation of the two' (Chambers & Trudgill 1980: 165), it seems reasonable to assume that the different patterns of synchronic variation are 'somehow a result of differences in the rate of linguistic change; differences probably resulting from social constraints on change.
An Integrated Approach to the Description of Linguistic Variation at the Cape

Statistics can only tell us what is normal. What is natural - the goal of linguistic theory - must be determined on the basis of patterns.

C.-J. Bailey (1980)

There are not many differences in mental habit more significant than that between the habit of thinking in discrete, well-defined class concepts and that of thinking in terms of continuity, of indefinitely delicate shadings-off of everything into something else, of the overlapping of essences.

Arthur O. Lovejoy The Great Chain of Being (1936)

8.1 Identifying Lects in the Data

8.1.1 Multivariate Analysis of Ten Linguistic Variables

While in chapters 6 and 7 the distribution patterns of the individual linguistic variables were considered separately, the final (multivariate) analysis focuses on patterns of co-variation between the different variables. It was argued in chapter five that numerical clustering methods such as cluster analysis, multidimensional scaling and principal components analysis (PCA) are useful tools for the grouping of entities (in this case idiolects), and allow for the identification of large-scale linguistic varieties in terms of similarity and co-occurrence.

Ten linguistic variables, all of which are of diagnostic value for the description of Afrikaans as an independent linguistic system distinct from Dutch, were included in the multivariate analysis:

- loss of neuter gender, understood as loss of agreement between article/demonstrative and neuter noun (variable GENDER)
- loss of person and number distinction in the present tense, main verbs only (variable VERB)
- loss of inflection in the infinitive (variable INF)
- loss of the inflectional ending in the past participle, loss of -t and -en considered separately (variables PP_T and PP_EN)
- loss of the preterite and generalisation of the perfect, main verbs only (variable TENSE)
- use of the 'Afrikaans' pronoun variants, combined scores for the use of ons, hulle, dit, my/sy (poss) and relative wat (variable PRONOUN)
- generalisation of the om...te infinitive (variable OMTE)
- use of Nie-2 (variable NIE2)
- use of objective vir per 1000 words (variable VIR)

Except for VIR all other variables were quantified as percentages (number of occurrences/number of loci). The decision to use combined scores for the pronoun system was motivated by
pragmatic considerations, i.e. only few individuals had scores (according to the criteria outlined in § 4.2) on all the different pronominal subcategories.

The results from hierarchical cluster analysis (Ward's method and average linkage; distance measure: Euclidean) indicate that in terms of their linguistic performance three main linguistic groups can be distinguished (see figure 8.1 and 8.2 for the results from Ward's method).

Figure 8.1 Tree diagram (Ward's method, Euclidean distances) for the variables GENDER, VERB, INF, PP_T, PP_EN, TENSE, PRONOUN, OMTE, NIE2 and VIR. Scores on all ten variables acc. to the criteria outlined in § 4.2 were available for 37 cases/individuals.

Figure 8.2 Graph of amalgamation schedule for Ward's method.

The three-cluster solution was repeated using k-means clustering, and cluster membership was found to be stable with four exceptions (A. Hauptfleisch, C. Leipoldt, M. Basson and L. Botha). Using multidimensional scaling the relative location of the individuals in the two-
dimensional space was computed (figure 8.3; the Shepard's diagram indicating the goodness of fit of the multidimensional scaling plot is given in figure 8.4). Following Berruto (1987: 265) who described linguistic varieties as 'Verdichtungspunkte in einem Kontinuum,' the three clusters can be seen as indicating the existence of three linguistic codes or varieties.

Figure 8.3 Multidimensional scaling (Guttman Lingoes) of the distance matrix for cluster I-III (GENDER, VERB, INF, PP_T, PP_EN, TENSE, PRONOUN, OMTE, NIE2 and VIR); based on Euclidean distances. 40 iterations computed, alienation =0.04, stress =0.04. Cluster I = 19 cases, cluster II = 8 cases, cluster III = 10 cases. Individuals with multiple cluster membership are indicated by a solid circle.

Figure 8.4 Shepard's diagram.
Dimension I in the multidimensional scaling plot (figure 8.3) can be interpreted as representing the progression from Dutch to Afrikaans, that is, individuals located furthest to the left (cluster I) have largely maintained the Dutch system, while the linguistic behaviour of individuals furthest to the right (cluster III) conforms closely to the system of Afrikaans. This interpretation was confirmed using multiple regression analysis. High correlation coefficients (ideally around 0.9) can be interpreted as relating systematically to position in the two-dimensional configuration (table 8.1).

<table>
<thead>
<tr>
<th></th>
<th>Dimension I</th>
<th>Dimension II</th>
</tr>
</thead>
<tbody>
<tr>
<td>GENDER</td>
<td>0.942</td>
<td>-0.08</td>
</tr>
<tr>
<td>VERB</td>
<td>0.962</td>
<td>0.033</td>
</tr>
<tr>
<td>INF</td>
<td>0.987</td>
<td>0.013</td>
</tr>
<tr>
<td>PP_T</td>
<td>0.970</td>
<td>-0.01</td>
</tr>
<tr>
<td>PP_EN</td>
<td>0.916</td>
<td>0.277</td>
</tr>
<tr>
<td>TENSE</td>
<td>0.705</td>
<td>0.295</td>
</tr>
<tr>
<td>PRONOUN</td>
<td>0.951</td>
<td>-0.03</td>
</tr>
<tr>
<td>OMTE</td>
<td>0.651</td>
<td>-0.42</td>
</tr>
<tr>
<td>NIE-2</td>
<td>0.842</td>
<td>-0.20</td>
</tr>
<tr>
<td>VIR</td>
<td>0.687</td>
<td>-0.19</td>
</tr>
</tbody>
</table>

Table 8.1 Multiple regression (beta correlation coefficient). Correlations are significant at p < 0.01.

All ten variables correlate highly and positively with dimension I. Dimension II is characterised by two weak correlations (TENSE, PP_EN), and one moderate (negative) correlation on the variable OMTE. In other words, individuals with high scores on dimension I show high scores on all ten variables; individuals with high scores on dimension II show relatively low scores on variable OMTE, and moderately high scores on TENSE and PP_EN (for a discussion of correlation coefficient values and their interpretation see Rowntree 1991: 169).

The linguistic behaviour of the three clusters can be summarised by plotting the cluster mean for each variable (figure 8.4). The structure of the plot of means indicates that the three clusters show what Labov has termed a 'complex pattern of regular relations' (1991[1972]: 124), i.e. they differ in their linguistic behaviour quantitatively, but the overall pattern is symmetrical.

---

1 The use of multiple regression analysis as an aid for the interpretation of the multidimensional scaling plot was suggested by Kruskal & Wish (1978). Briefly, the idea is to use the coordinates of the configuration as the independent variable, and the linguistic variables as the dependent variable. Multiple regression analysis is a parametric technique used to analyse the relationship (correlation) between two or more independent variables and a single dependent variable. It assumes that the relationship between dependent and independent variables is linear and that the variables are normally distributed. While the inspection of the individual scatterplots showed that the variables in question did not show curvature in their relationships, the variables were not normally distributed (the normality assumption was tested using the Shapiro Wilks W test). Although multiple regression is quite robust to violations of the normality assumption, the non-parametric method Spearman R was used to confirm the multiple regression results given in table 8.1.
This is most clearly visible for the morphosyntactic variables where all three groups show roughly the following hierarchy for the frequency pattern of morphosyntactic reduction:

$$\text{GENDER} \geq \text{VERB} > \text{PP\_T} \geq \text{INF} > \text{PP\_EN}$$

A gradual pattern is also visible for the generalisation of the perfect and the om...te infinitive. Here high percentages are not only found in clusters II and III, but also in cluster I whose members conform closely to the system of continental Dutch with regard to the morphosyntactic and morpholexical variables. This supports the item-based analysis presented in § 6.2.1 and § 7.3.2 which showed that both variables were already widely diffused through the Cape speech community. For the pronoun system and the brace negation, only cluster III shows high scores, while the scores of both cluster I and II lie well below 30%.

![Figure 8.5 Plot of means for the three cluster solutions (GENDER, VERB, INF, PP\_T, PP\_EN, TENSE, PRONOUN, OMTE, NIE2 and VIR, based on the results of k-means clustering); C= cluster.](image-url)
The clusters identified by the clustering algorithm must be understood as descriptive constructs which do not necessarily have a 'reality' in the minds of speakers or 'the world' (McEntegart & LePage 1982: 115-116, see also § 5.2.2). The visual representation of the cluster solution in figure 8.3 suggests that the identification of cluster III fulfills one of the criteria usually cited as typical for 'natural' classifications, i.e. cluster III is well separated in the two-dimensional space. However, the variable cluster membership of four individuals (A. Hauptfleisch, M. Basson, C. Leipoldt and L. Botha) indicates that the boundary between cluster I and II is rather ill-defined, and these individuals can be said to have degrees of membership in both clusters. Such examples of case instability as well as the general dispersion of cases within all three clusters suggests that the clusters show properties of fuzzy sets: they constitute sets characterised by graded membership and ill-defined boundaries (see Zadeh 1965, also McNeill & Freiberger 1993).  

Turning to the degree of within-cluster variance (another diagnostic criterion for naturalness, see § 5.2.2.), the statistical measure of 'variance' provides a useful index of the degree to which the individual scores within a cluster deviate from the cluster mean. A small variance indicates that most of the scores lie close to the mean, a large variance represents scores that are widely scattered.

---

Table 8.2 Cluster means for all ten variables as well as the corpus mean (all in %). Note that the mean percentage given for the corpus in this table is not based on the actual number of tokens, but calculated on the basis of the percentages given for each individual, i.e. only individuals with a minimum of six tokens were included.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Cluster I</th>
<th>Cluster II</th>
<th>Cluster III</th>
<th>Corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>15</td>
<td>70</td>
<td>98</td>
<td>51</td>
</tr>
<tr>
<td>Verb</td>
<td>7</td>
<td>70</td>
<td>96</td>
<td>46</td>
</tr>
<tr>
<td>Inf</td>
<td>3</td>
<td>51</td>
<td>88</td>
<td>38</td>
</tr>
<tr>
<td>PP_T</td>
<td>3</td>
<td>60</td>
<td>90</td>
<td>42</td>
</tr>
<tr>
<td>PP_EN</td>
<td>1</td>
<td>44</td>
<td>78</td>
<td>29</td>
</tr>
<tr>
<td>Tense</td>
<td>65</td>
<td>90</td>
<td>98</td>
<td>78</td>
</tr>
<tr>
<td>Pronoun</td>
<td>3</td>
<td>25</td>
<td>92</td>
<td>28</td>
</tr>
<tr>
<td>Omte</td>
<td>37</td>
<td>48</td>
<td>75</td>
<td>48</td>
</tr>
<tr>
<td>NIE2</td>
<td>0</td>
<td>5</td>
<td>68</td>
<td>16</td>
</tr>
<tr>
<td>Vir</td>
<td>0.2</td>
<td>2</td>
<td>5</td>
<td>2.0</td>
</tr>
</tbody>
</table>

2 The percentages based on the total number of tokens in the corpus are: Gender 57%, Verb 58%, Inf 40%, PP_T 45%, PP_EN 32%, Tense 75%, Pronoun 35%, Omte 52%, NIE2 21% and Vir: 2 per 1000 words.

3 Unfortunately STATISTICA does not allow for 'fuzzy' cluster analysis. Kaufman & Rousseeuw (1990: 42) mention a programme called FANNY which employs a clustering algorithm based on fuzzy set theory, i.e. rather than assigning a case categorically to a given cluster FANNY can indicate, for example, that a case X belongs 80% to cluster I and 20% to cluster II. Fuzzy clustering is also available for SAS (see SAS/ SAT User's Guide 1988: 47-83; for a discussion of fuzzy clustering see Zimmerman 1991: 220-239).
The information presented in table 8.3 shows that the three clusters can be ordered in terms of variance: cluster II > cluster III > cluster I. The individual variance values indicate that the different variables contribute to differing degrees to the variation within the data set as a whole as well as within clusters. Considering the entire corpus, the variance values indicate that the loss of the preterite tense (TENSE), the generalisation of the om...te infinitive (OMTE), and the double negation (NIE2) contribute least to the variation in the data. For the variables TENSE and OMTE this is due to their relatively broad diffusion; for the variable NIE2, lack of use by the majority of individuals.

Cluster I is best described as acrolectal, i.e. the individuals grouped in this cluster conform relatively closely to the norms of continental Dutch. Variation is most pronounced on the variables OMTE and TENSE, while variation in the strong past participle which is substantial in the other two clusters, is marginal. There is also considerable variation on the variables GENDER and VERB, which indicates that within limits morphological reduction had also affected acrolectal varieties. That is, while acrolectal varieties generally maintained the distinction between finite and infinite verbs as well as the inflectional marking of the past participle, they increasingly showed loss of gender and loss of person and number distinction of the verb. It should be noted in this context that the measure of variance can be misleading in cases where the data set contains so-called outliers or residuals (i.e. single unusual data values) which can unduly influence the size of the mean (i.e. the basis on which variance in calculated). In descriptive and exploratory work alternative measures of central tendency which are more resistant to outliers (mode and median and 'their' measures of dispersion, i.e. range and interquartile range), are thus often seen as preferable to the more conventional measures of mean and standard deviation (see Marsh 1988: 34-36). With respect to cluster I, median and interquartile range indicate that within
this cluster outliers exist for the variables OMTE and TENSE, thus, giving the impression that the cluster is relatively variable; however the variability is largely an artefact resulting from the existence of few unusual data values. This interpretation is supported the visual representation of the cluster (figure 8.3) in which a cluster core (i.e. several closely situated data points) and outliers are noticeable.

As visible in the scatterplot and confirmed by the variance values, the intermediate cluster II is extremely variable. Again, the degree of within-cluster variation is not evenly distributed across the different variables. Variation is particularly prominent for the variable PP_EN, followed by PP_T, OMTE, TENSE and GENDER. Variation of NIE2 is low as only one group member (H.S. Vermaak) uses the double negation at all. Individuals grouped together in this cluster are not only somehow 'intermediate', but their linguistic usage positions them either closer to the new norms of Afrikaans (cluster III) or to the old continental standard of Dutch (cluster I). The broad dispersion of cases within this cluster, as well as the existence of a fuzzy boundary with cluster I, suggests the possibility of a pattern of continuous variation bridging the gap between the poles from Dutch to Afrikaans.4

Like Cluster I, Cluster III has comparatively low variance values and can be interpreted as a focused variety in the sense in which LePage and Tabouret-Keller (1985: 181-182) use the term. As noted in the introduction, linguistic focusing refers to the gradual reduction of variation and relates to processes of standardisation, and Cluster III might, thus, be classified as an emerging standard. Variation in cluster III is notable only for loss of -e(n) in the past participle (i.e. several individuals grouped in this cluster tend to maintain the inflectional ending for strong past participles) and nie-2. The generalisation of the om...te infinitive is also still variable, with some individuals still making frequent use of the te infinitive.

Comparing clusters I (the acrolectal Dutch standard) and III (the emerging standard of Afrikaans), one can argue that what we observe here is the gradual disintegration of the old standard (giving rise to an increasingly diffuse set of acrolectal norms), and the equally gradual focusing of the new standard (indicated by decreasing variability). The visual representation in the scatterplot (figure 8.3) shows cluster I as consisting of a cluster core surrounded by peripheral cluster members (outliers; the cluster is tight and dense, see § 5.2.2). Cluster III, on the other hand, does not yet show a cluster core, however, the data points (although dispersed) are comparatively close together (i.e. the cluster is loose but shows high density). It can be

4 Note that the tree diagram (figure 8.1) and the graph of amalgamation schedule (figure 8.2) indicate the possibility of distinguishing five rather than three groups in the data. According to the results from both Ward's method and k-means clustering this would result in separating cluster I into a core and a periphery, and cluster II would be broken off into two clusters (one closer to cluster I and one closer to cluster III; these could be described as upper/ lower mesolectal). However, as especially the graph of amalgamation schedule clearly levels off after step three, the three cluster solution was seen as preferable.
hypothesised that disintegration will continue in cluster I, while the data points in the emerging standard of cluster III will gradually move closer together. The processes of disintegration and focusing are illustrate in figures 8.6 and 8.7.

Figure 8.6 The disintegration of an existing standard variety: diffusion. The arrow indicates the direction of the development.

Figure 8.7 The emergence of a new standard variety: focusing. The arrow indicates the direction of the development.

Returning to the notion of co-occurrence as a diagnostic criterion for the identification of large-scale linguistic varieties (see § 5.3), the correlation matrix given in table 8.4 shows that most of the ten variables are highly and positively correlated, i.e. high (low) values on variable VERB go with high (low) values on variable INF, variable PP_T, and so forth. More moderate correlations exist for the variables TENSE, OMTE and also VIR and NIE2.
Table 8.4 Correlation matrix for the ten variables. Pearson \( r \); correlations are significant at \( p<0.05 \).

The existence of strong correlations between the variables was confirmed by the results from principal components analysis. Based on the interpretation of the scree plot (figure 8.8), two factors with eigenvalues of 7.7 and 0.6 were maintained. Together the two factors account for 82.9% of the variance in the data.

The first factor which is characterised by high positive loadings on all ten variables, accounts for 76.5% of the variance. The second factor is only weakly characterised by a relatively high, negative loading on OMTE, and a moderate, positive loading on TENSE. Thus, again the two variables, which are most widely diffused in the sample, are shown to occupy a special position.
The weak loadings on factor II as well as the structure of the scree plot (figure 8.7) suggest that variation is largely uni-dimensional. In other words, from cluster I to cluster III there is a linear increase in the frequency of 'Afrikaans' forms on all ten variables.5

The analysis of the individual variables described in chapters six and seven has indicated that the morphosyntactic variables differed from the morpholexical variables with regard to their distribution across cases. That is, in the case distribution for morphosyntactic variables continuity was found to interact with polarisation, while the case distribution for the morpholexical variables was dichotomous with few individuals in the middle field (§7.4). To investigate these differences in more detail, cluster analysis and multidimensional scaling were performed, first only including the morphosyntactic variables (loss of verbal inflection, loss of gender), and then including only the pronoun system.6

Morphosyntactic Variation: Considering the verbal system (variables VERB, INF, PP_T, PP_EN)

---

5 McEntegart & LePage (1982: 120-212) remarked that highly correlated variables are essentially reporting the same information, and could thus have an undue effect on the clustering results: 'The fear is that in calculating similarities of the observations over such a set of variables, a form of double counting takes place. The true nature of the similarity/distance between the individuals may be obscured by this double counting.' (1982: 120). However, to discover that variables are highly correlated is in itself an important result, rather than a mere 'problem' leading to data distortion.

6 Considering morphosyntactic and morpholexical variation separately is of interest also when considering the recent discussions of modularity in linguistics, i.e. the idea that parts of the grammar form sub-systems, that different sub-systems might behave differently, and thus have some kind of analytical independence (see Ferguson 1995: 191-192).
separately, a three cluster solution was again indicated by the different graphs of amalgamation schedule (Ward's method, and average linkage; distance measure: Euclidean); however, the structure of the solution differs in important ways from the results discussed in § 8.1.1. The intermediate cluster II shifts to the right towards the emerging standard of Afrikaans (cluster III), and the fact that the cluster membership of four individuals (E. Furter, E. Smith, E. Steyn and J. Brümmer [2]) was unstable across methods suggests increasing fuzziness between cluster II and III. The gap between cluster I and II, on the other hand, is slightly more pronounced, and only one individual (A. Hauptfleisch [1]) shows variable cluster membership. The means for the three clusters again indicate a symmetric hierarchy, i.e. VERB > PP_T > INF > PP_EN.

Figure 8.9 Multidimensional scaling (Guttman Lingoes) of the distances matrix for cluster I-III (VERB, INF, PP_T, PP_EN); based on Euclidean distances. 49 iterations computed; alienation 0.03, stress 0.03. Cluster I= 24 cases, cluster II= 11 cases, cluster III= 15 cases. Individuals with variable cluster membership are indicated by a solid circle.

The general fuzziness of the solution suggests the possibility that in this case the clustering algorithm 'imposed' a structure on a situation which is probably more accurately described as a continuum, i.e. a finely graded 'spectrum of usage' (DeCamp 1971b: 36) which is 'not amenable to either, mono- or bisystemic treatment' (Bickerton 1973: 641). DeCamp (1971a, 1971b) has pointed to the implicational relations which often characterise linguistic continua, and following his analysis of seven Jamaican speech samples Guttman scalogram analysis is today commonly used for the description of implicational continua (for the methodology see Guttman
1944, Togerson 1958: 307-331). Early implicational scales were generally based on binary data (i.e. a certain linguistic feature was classified as being either absent or present); however, as linguistic data usually do not come in such an all-or-nothing way, Bickerton (1973: 642, see also Fasold 1970: 553-554) suggested the use of three-place scales which allow for the incorporation of variable structures. A 'perfect' continuum using a three-place-scale is illustrated in table 8.6. The implicational relations which exist in the continuum can be summarised as follows: the presence of property D in the speech of an individual means that the speaker will also use properties A to C; if D is absent but C is present then properties A and B will also be present, and so forth. In other words, D implies C implies B implies A. Perfect continua are, however, not to be expected in empirical research, but rather constitute theoretical models which describe the structure of linguistic variability in speech communities. Thus, DeCamp (1971b: 35) likened the continuum model to the perfect shapes which underlie the study of geometry.

You do not obtain a square by carefully measuring thousands of floor tiles and then averaging the measurements, for a square has four equal sides by a priori definition, not by empirical measurement. However, geometry is useful to us precisely because there are shapes in our real world which, though irregular, are similar enough to the ideal geometrical shapes that they can be described by surveyors and navigators as if they really were circles, squares, and triangles.

<table>
<thead>
<tr>
<th>ENVIRONMENT</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>lect 1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>lect 2</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>V</td>
</tr>
<tr>
<td>lect 3</td>
<td>X</td>
<td>X</td>
<td>V</td>
<td>V</td>
</tr>
<tr>
<td>lect 4</td>
<td>X</td>
<td>V</td>
<td>V</td>
<td>V</td>
</tr>
<tr>
<td>lect 5</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>V</td>
</tr>
<tr>
<td>lect 6</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>O</td>
</tr>
<tr>
<td>lect 7</td>
<td>V</td>
<td>V</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>lect 8</td>
<td>V</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>lect 9</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

Table 8.6 Perfect implicational scale (adapted from Romaine 1982a: 171).

The systematic ordering of idiolects portrayed in implicational scales has been interpreted as a 'synchronic [reflection] of diachronic processes, depicting the acquisition or diffusion of features as waves spreading across social (speaker-to-speaker) and linguistic (environment-to-environment) space' (Rickford 1987b: 155, also Bailey 1973: 67-109). In other words, implicational scales are believed to illustrate the path of diachronic change from one system to another, and to preserve 'diachronic changes in a synchronic state' (Bickerton 1975: 17). The data for the variables VERB, INF, PP_T and PP_EN were scaled using a three-place scale with the following thresholds:
O 0-25%, i.e. largely corresponding to the norms of continental Dutch

V 26-74%, i.e. variable usage

X 75-100%, i.e. largely corresponding to the emerging norms of Afrikaans

Following DeCamp (1971b: 34) these three categories are (at least in some sense) meant 'to model the idealized competence of the persons involved in those speech acts'. It is of course questionable whether the setting of such ultimately arbitrary thresholds is theoretically acceptable (see Romaine 1982a: 178, fn. 19); however, relying on similarly arbitrary thresholds Romaine (1982a: 170-173) used implicational scaling successfully to illustrate the syntactic diffusion of relative deletion in Scots English.7

Table 8.7 shows the scaling of the data for 51 individuals. A key concept in scalogram analysis is reproducibility, since in a perfect scale the linguistic behaviour of any individual can be reproduced from the knowledge of its rank position alone. However, perfect scales are not to be expected in actual research and a coefficient of reproducibility (CR) is calculated to measure the degree of deviation from a perfect scale (CR = 1 - \( \sum \) errors/ \( \sum \) possible errors). Scales with a CR > 0.85 can be considered as approximations to a perfect scale, and a CR \( \geq \) 0.93 approaches the 0.05 level of significance (Guttman 1944: 140, Dunn-Rakin 1983: 105-107). Note, in table 8.7 empty cells were included as deviations for the calculation of the CR, a procedure which is not usually followed by linguists but was suggested by Romaine (1982a: 176).

<table>
<thead>
<tr>
<th>N=8</th>
<th>V</th>
<th>X</th>
<th>X</th>
<th>X</th>
<th>X</th>
<th>AFRIKAANS</th>
</tr>
</thead>
<tbody>
<tr>
<td>N=1</td>
<td>V</td>
<td>X</td>
<td>X</td>
<td>X*</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>N=1</td>
<td>V</td>
<td>X</td>
<td>V</td>
<td>X</td>
<td>V</td>
<td></td>
</tr>
<tr>
<td>N=6</td>
<td>X</td>
<td>X</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td></td>
</tr>
<tr>
<td>N=3</td>
<td>X</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>O*</td>
<td></td>
</tr>
<tr>
<td>N=3</td>
<td>V</td>
<td>X</td>
<td>V</td>
<td>V</td>
<td>X*</td>
<td></td>
</tr>
<tr>
<td>N=1</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td></td>
</tr>
<tr>
<td>N=2</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>O</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=1</td>
<td>V</td>
<td>V</td>
<td>O*</td>
<td>V</td>
<td>O</td>
<td></td>
</tr>
<tr>
<td>N=0</td>
<td>V</td>
<td>V</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td></td>
</tr>
<tr>
<td>N=4</td>
<td>V</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=19</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 8.7 Implicational analysis including the variables VERB, INF, PP_T, PP_EN (for 50 cases). Deviations indicated by *, CR=0.93.

7 Romaine (1982a: 172) specified the following thresholds for her analysis of relative deletion in Middle Scots: O= 5% or less deletion, X= 5-25% deletion, 1 = 25% or more deletion.
The implicational analysis supports the hierarchical ordering of the four verbal variables established on the basis of cluster means (see § 8.1.1). In other words, the hierarchical ordering
\[
\text{VERB} > \text{PP\_T} > \text{INF} > \text{PP\_EN}
\]
is not an artefact of group analysis but also reflected in the individual lects, possibly indicating the diachronic direction of the diffusion of morphological simplification through grammatical categories. The fact that implicational relations are shown to exist in the data indicates that linguistic variation was structured and systematic, and that the intermediate lects as grouped in cluster II are not the result of random dialect mixture (under the influence of the standard Dutch) but linguistic systems in their own right. The scale shows, furthermore, that morphosyntactic variation at the Cape was finely graded, and is thus best described by using the theoretical construct of a continuum.

The implicational relations hold when including loss of gender as a fifth morphosyntactic variable (table 8.8). The CR is, however, lower when counting empty cells as deviations (CR = 0.86). However, since the scale is based on fewer individuals (N=40), it is possible that the empty cells would be filled if more data were available (see Rickford 1987a: 20). Calculating the CR on the basis of deviant cells alone approximates the 0.05 level of significance (CR=0.95).

<table>
<thead>
<tr>
<th>GENDER</th>
<th>VERB</th>
<th>PP_T</th>
<th>INF</th>
<th>PP_EN</th>
</tr>
</thead>
<tbody>
<tr>
<td>N=7</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>N=4</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>V</td>
</tr>
<tr>
<td>N=1</td>
<td>X</td>
<td>X</td>
<td>V</td>
<td>V</td>
</tr>
<tr>
<td>N=2</td>
<td>X*</td>
<td>X</td>
<td>V</td>
<td>V</td>
</tr>
<tr>
<td>N=1</td>
<td>V</td>
<td>X</td>
<td>V</td>
<td>X*</td>
</tr>
<tr>
<td>N=0</td>
<td>X</td>
<td>V</td>
<td>V</td>
<td>V</td>
</tr>
<tr>
<td>N=1</td>
<td>X</td>
<td>V</td>
<td>V</td>
<td>X*</td>
</tr>
<tr>
<td>N=0</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>V</td>
</tr>
<tr>
<td>N=2</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>V</td>
</tr>
<tr>
<td>N=1</td>
<td>V</td>
<td>X*</td>
<td>V</td>
<td>V</td>
</tr>
<tr>
<td>N=0</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>V</td>
</tr>
<tr>
<td>N=1</td>
<td>V</td>
<td>V</td>
<td>V*</td>
<td>V</td>
</tr>
<tr>
<td>N=0</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>V</td>
</tr>
<tr>
<td>N=3</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>V</td>
</tr>
<tr>
<td>N=1</td>
<td>O</td>
<td>V*</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>N=14</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

Table 8.8. Implicational analysis including the variables GENDER, VERB, INF, PP\_T AND PP\_EN (for 41 cases). Deviations indicated by *, CR (empty cells counted as deviant) = 0.85, CR (empty cells excluded)= 0.95.

Morpholexical Variation: Multivariate analysis was performed using the pronoun variables ONS, DIT, MY/SY (possessive) and HULLE (subject and object combined; scores on all four pronoun variables available for 20 individuals). The results differ in important ways from those discussed
for the morphosyntactic variables. Cluster analysis indicates an unambiguous two cluster solution (distance measure Euclidean, linkage rules: Ward's method, complete and average linkage), which was repeated using k-means clustering. Cluster membership was entirely stable.

![Multidimensional scaling](image)

Figure 8.10 Multidimensional scaling (Guttman-Lingoes) of the distance matrix for cluster I and II (ONS, DIT, MY/SY, HULLE); Euclidean distances. 173 iterations computed, alienation = 0.01, stress 0.008. Cluster I= 12 cases, cluster II= 8 cases.

Unlike the previous cases, no intermediate varieties are indicated and the distribution is unambiguously bi-systemic. The bi-systemic interpretation of the solution is supported by the clear gap between the two groups as well as the density of the two clusters. Dimension I can easily be identified as leading from Dutch (cluster I) to the new standard of Afrikaans (cluster II). Dimension II is necessary only for the placement of F.S. Malan, who conforms to the new standard of Afrikaans with one exception: he still uses the possessive pronoun ending in a nasal (myn/syn). Individuals who were previously classified as belonging to cluster II are now grouped as (peripheral) members of cluster I. In other words, membership of cluster III remained stable, and cluster II collapsed with cluster I.

---

The variables EK and HOM were included in a control analysis which was conducted to ensure the stability of the solution, and the two cluster solution was repeated. The two clusters which were again clearly separated in the two-dimensional space, showed more within-cluster-variability. This is not surprising since EK and HOM were less diffused than the other pronoun variables (see § 7.1.1).
Summary: The results of the multivariate analysis can be summarised as follows:

- All ten variables are highly and positively correlated, and the results from PCA suggest that over two-thirds of the variation in the data can be summarised by one single, linear factor which can be described as [±Afrikaans].

- Morphological reduction (verbal and nominal) is continuous, implicational and linearly increasing from cluster I to III.

- Variation on the variables TENSE and OMTN also increases linearly from cluster I to III; however, the frequency differences between the three clusters are much less pronounced than on the other variables, indicating that by the turn of the century change on these two variables had affected most codes/lects used within the Netherlandic speech community. This supports the results from the item-based analysis in § 6.2.1 and § 7.2.2.

- Variation on the morpholexical variables (pronoun system) is bi-systemic. It is possible that this reflects the existence of multiple norms in the spoken language.

Considering the distribution of the three clusters in the two-dimensional space, one might want to describe them with reference to the terminology of acrolect, mesolect and basilect (see § 2.3). However, it will be shown in the next section that the social composition of cluster III is at odds with the social characteristics typically reported for basilectal varieties.

8.1.2 Examining the Social Dimensions

So far the discussion has focused on the patterns of linguistic variation in the data set and ignored the social location of the speakers. While general correlation analysis (using multiple regression) was inconclusive, an interesting social pattern was found to exist in the three cluster solution discussed in § 8.1.1. Of the three clusters, no cluster is entirely homogenous with regard to social criteria; however, cluster membership is not random, but clearly structured along social dimensions (table 8.9).

Unfortunately the actual numbers in table 8.9 are marginal, and since the reliability of the $\chi^2$ test depends on sufficiently large expected frequencies (generally 5 or greater), testing for statistical significance of the distribution of social characteristics is problematic. Using the $\chi^2$-test under conditions of low expected frequencies increases the likelihood of what is known as a type 1 error ($\alpha$-error), i.e. the rejection of the null-hypothesis (that is, the assumption that the distribution in the sample does not exist in the population from which the sample is drawn) when
it is true (Woods, Fletcher & Hughes 1986: 145). At the same time type 2 errors (β error) (rejection of the alternative hypothesis when it is true) are generally likely when the sample size is very small (ibid.: 127, 130). The results of the $\chi^2$ test are thus tentative, and should be seen as indicators of trends or tendencies, rather than allowing for firm conclusions.

<table>
<thead>
<tr>
<th></th>
<th>Sex</th>
<th>Age Group</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>M</td>
<td>1</td>
</tr>
<tr>
<td>All</td>
<td>19</td>
<td>18</td>
<td>13</td>
</tr>
<tr>
<td>%</td>
<td>51%</td>
<td>49%</td>
<td>35%</td>
</tr>
<tr>
<td>Cluster I</td>
<td>9</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>%</td>
<td>47%</td>
<td>53%</td>
<td>38%</td>
</tr>
<tr>
<td>Cluster II</td>
<td>5</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>%</td>
<td>63%</td>
<td>37%</td>
<td>63%</td>
</tr>
<tr>
<td>Cluster III</td>
<td>5</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>%</td>
<td>50%</td>
<td>50%</td>
<td>10%</td>
</tr>
</tbody>
</table>

Table 8.9 Social characteristics of the three cluster solution (37 individuals, ten linguistic variables).

Cluster III (the emerging standard of Afrikans) is characterised by a dominance of age group 2 ($\chi^2 = 2.747253$, df=9, $p=0.09$), and a smaller but still considerable dominance of members of the new professional class ($\chi^2 = 3.6000$, df=1, $p=0.06$). Note that the class 3 individuals included in cluster III are all young (age group 2) and relatively well educated. All of them were classified as class 3b (see § 4.3.2), i.e. they were independent small-scale entrepreneurs with at least moderate education (i.e. several years of schooling). The intermediate cluster II is entirely composed of members of class 3 (almost exclusively class 3a; $\chi^2 = 8.0$, df=1, $p=0.01$). The cluster also contains a relatively high percentage of older speakers (age group 1, $\chi^2 = 2.133333$, df=1, $p=0.1$), especially when compared to their under-representation in the split sample as a whole. Note that commentators in the 19th century frequently remarked on the existence of generational differences, and described the language of the older generation as being 'more Dutch' (see § 3.4). With regard to sex, cluster II is characterised by a slight predominance of women, and considering that this in-between cluster is both progressive and conservative, the distribution agrees with Raidt's (1995: 138) assessment of the role of women in linguistic change at the Cape: 'women in the Cape Colony played a double role in linguistic change - both progressive and conservative'. Finally, cluster I (the old Dutch standard) does not show any particular bias, but reflects roughly the general social distribution of the split sample.

The overall social patterning across clusters I to III holds when using only subsets of the linguistic variables, such the as the verbal variables ($\text{VERB}, \text{INF}, \text{PP_T}$ and $\text{PP_EN}$) discussed in the previous section. However, additional tendencies are noticeable in the social distribution.
The social patterning across the three subgroups (established on the basis of the morphosyntactic variables only) is summarised in table 8.10.

<table>
<thead>
<tr>
<th>Sex</th>
<th>Age Group</th>
<th>Class</th>
<th>Ethnicity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>M</td>
<td>1</td>
</tr>
<tr>
<td>All</td>
<td>N</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>Cluster I</td>
<td>N</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>46%</td>
<td>54%</td>
</tr>
<tr>
<td>Cluster II</td>
<td>N</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>45%</td>
<td>55%</td>
</tr>
<tr>
<td>Cluster III</td>
<td>N</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>60%</td>
<td>40%</td>
</tr>
</tbody>
</table>

Table 8.10 Social characteristics of the three cluster solution for morphosyntactic variables (based on four linguistic variables VERB, INF, PP_T and PP_EN, 50 individuals).

The intermediate cluster II is again characterised by dominance of the oldest age group and of class 3. The higher percentage of class 3 and age group 1 in cluster III is a result of excluding the more dichotomously patterned variables PRONOUN and NIE2, and re-introducing these variables into the analysis immediately reduced the percentage of class 3/age group 1 in cluster III (i.e. these individuals then move 'back' into cluster II). Thus, the morphosyntactic variables which showed continuity across cases, are also less categorically structured on the social dimensions. Note that except for the variable ethnicity, cluster I (the old standard of Dutch) and cluster III (the new standard of Afrikaans) show a very similar social distribution. Regarding the role of ethnicity the distribution is, however, inconclusive because of the marginal numbers involved.

Turning to the more dichotomously patterned morpholexical variables, the social correlates of the two pronoun clusters discussed above are summarised in table 8.11.

<table>
<thead>
<tr>
<th>Sex</th>
<th>Age</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>M</td>
</tr>
<tr>
<td>All</td>
<td>N</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>60%</td>
</tr>
<tr>
<td>Cluster I</td>
<td>N</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>58%</td>
</tr>
<tr>
<td>Cluster III</td>
<td>N</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>63%</td>
</tr>
</tbody>
</table>

Table 8.11 Social characteristics of the two cluster solution for morpholexical variables (based on four linguistic variables ONS, DIT, HULLE and MY/SY (possessive), 20 individuals).

The data given in table 8.11 supports the interpretation of the two-cluster solution as
eliminating the intermediate cluster II: individuals who with respect to their use of verbal inflections and nominal gender marking, fall into the intermediate group or even cluster III, move into cluster I (acrolect), thus changing its social composition (i.e. increasing the percentage of class 3 and age group 1).

In sum, although the social correlates of variation across the three clusters are not categorical, and to some extent dependent on the linguistic variables involved, the overall tendency is stable. That is, while the linguistic patterns across the clusters are linear (i.e. gradual increase in the frequency of 'Afrikaans' forms, see figure 8.11 and the discussion in § 8.1.1), the social patterns across the clusters are curvilinear with respect to age and class, as well as sex. The different patterns of relationships are summarised in figure 8.11 to 8.14.

![Figure 8.11 Distribution of 'New' or 'Afrikaans' forms across the three clusters (averaged across the ten variables, in percentages).](image1)

![Figure 8.12 Distribution of social groups across clusters: age (in percentages, from table 8.9).](image2)
Figure 8.13 Distribution of social groups across clusters: class (in percentages, from table 8.9).

Figure 8.14 Distribution of social groups across clusters: sex (in percentages, from table 8.9).

The three clusters can be described sociolinguistically as follows:

Cluster I  acrolectal Dutch, relatively low level of linguistic variation; dominated by age group 2, both class 2 and 3
[-Afrikaans, - variation, +age group 2, + class 2, + class 3]

Cluster II  mesolectal Dutch/Afrikaans, i.e. no clear systemic identity; high levels of within-cluster variation, dominated by age group 1 and class 3
[± Afrikaans, +variation, -age group 2, -class 2, +class 3, +female]
Cluster III: 'Afrikaans'; relatively low degrees of linguistic variation; dominated by age group 1 and class 2
[+Afrikaans, -variation, +age group 2, -class 2, -class 3]

As argued in § 2.3, the terminology of acro-, meso- and basilect has proved useful for the description of linguistic variation in extraterritorial settings. However, while most linguists would comfortably identify cluster I as acrolectal (i.e. more or less corresponding to the colonial language; probably best described as an ET dialect of standard Dutch), and cluster II as somehow mesolectal, labelling cluster III as basilectal does not seem appropriate since the concept of a basilect not only refers to a linguistic entity characterised by the fact that it is furthest removed from the acrolect (in terms of linguistic structure), but also to its social location and identity as a stigmatised low prestige variety. Considering the social composition of cluster III, especially the dominance of the new (highly educated) professional class, it appears more appropriate to label this cluster 'elevated basilect' or 'emerging non-conservative standard'. Thus, although situated at the opposing pole in terms of its linguistic identity, cluster III exhibits social characteristics typical for acrolectal varieties. That is, it is a relatively focused variety and is used by educated writers. Conspicuous by its absence from the corpus is what one might want to call the 'real colloquial', the 'true' basilect (see § 4.4).

8.1.3 Linguistic Patterns in the Vernacular Writing Tradition

It was suggested in § 3.1 that the vernacular writing tradition at the Cape might have contributed to the standardisation of Afrikaans by providing prospective writers with a typological conception of what constituted the 'vernacular'. To further investigate the structural (linguistic) location of the vernacular literature in the spectrum of variation at the Cape, a sample of nine vernacular texts (together over 18 000 words) was combined with the analysis of the split sample of 36 cases and ten linguistic variables (§ 8.1.1). The sample of vernacular texts covers the time period 1828 to 1909 and the following texts were included:

- Zamenspraak tusschen Klaas Waarzegger en Jan Twyfelaar, No 1 (author: L. Meurant 1861).
• Benigna van Groenekloof of Mambre (anonymous, 1873).
• Geskiedenis van Josef, chapter I-II (author: C.P. Hoogenhout, 1873).
• Manifestes van die Genootskap van Regte Afrikaners (AP 15/1/1876).
• Voor Volk en Taal, scene 5/6 (author: M. Brink, 1905).

Figure 8.15 shows the pictorial representation of the distribution of the 36 individuals and nine vernacular texts in the two-dimensional space.

![Graph showing distribution of texts](image-url)

Figure 8.15 Multidimensional scaling (Guttman Lingoes) of the distance matrix for cluster I-III (GENDER, VERB, INF, PP_T, PP_EN, TENSE, PRONOUN, OMTE, NIE2 and VIR for 37 cases and 9 vernacular texts); based on Euclidean distances. 42 iterations computed; alienation 0.05, stress 0.05. The location of individuals differs slightly from figure 8.3 as a result of the incorporation of further cases. Vernacular texts are indicated as follows: De Lima 1, Boniface 2, Meurant 3, Benigna 4, Zwaartman 5, Josef 6, Manifest 7, Brink 8, Straatpraatjes 9.

All vernacular texts except for those by De Lima and Brink are grouped with cluster III,

---

9 As the narrative of the novel is in Dutch, only those sections which reported direct speech were included into the analysis.
i.e. the emerging standard of Afrikaans. Nienaber (1971: 24) has described De Lima’s *Mietje*-letters as an example of what he termed 'Hoog-Afrikaans'; a variety presumably representative of the language used by the upper classes in Cape Town. From the perspective of the continental Dutch standard, De Lima’s language use is full of mistakes: gender marking of the noun is virtually absent (for example, *de huis, de werk, de ding*, but *het genoegge*), loss of verbal inflection is completed for modal verbs, but still variable for auxiliaries and main verbs as well as weak past participles. Infinitives and strong past participles are categorically inflected, and with respect to the more dichotomously patterned pronoun variables DeLima’s language still shows many characteristics of the system of Dutch. He uses only *wy* in the 1st person plural, *zulle* and *zy* for the 3rd person plural, and although he uses *dit* as a 3rd person singular pronoun, and *wat* as a relative (with non-animate antecedent only), examples of this are rare. No instances of Nie-2 exist in the *Mietje*-letters.

Brink, who explicitly aimed at writing what he called 'Cape Dutch' (see § 3.2), is further advanced towards Afrikaans than De Lima. His language use conforms to the Afrikaans system for auxiliary and modal verbs; however, he frequently marks inflection on main verbs (roughly one-third of his main verbs are inflected). Brink also shows a tendency to maintain inflection in the past participle. He has generalised the use of *ons* (instead of *wy*) and *dit* (instead of *het*), but shows variation for the 3rd person plural pronoun (*hulle~zulle*). Examples of Nie-2 occur but are infrequent (22% of negative sentences show Nie-2). Of the vernacular texts grouped in cluster III, Benigna and Boniface appear to be more peripheral; they show a still variable pronoun system, variation of Nie2 and in the past participle.

In sum, incorporating the sample of vernacular texts into the analysis gives empirical support to some hypotheses formulated in earlier chapters:

- The vernacular literature was typologically similar to what later was propagated as the Afrikaans (written) standard. Keeping in mind what we know about networks and exposure it is possible that the vernacular writing tradition contributed to the standardisation of Afrikaans by offering prospective writers a model of 'Afrikaans.'

- The linguistic similarities between the texts belonging to the vernacular writing tradition and the private letters of young and educated writers could be described with reference to what Britto (1986: 232) has termed a Pseudo-H. That is, a relatively stable linguistic variety typically found in eye dialects, where it is clearly identifiable as L (thus commanding local prestige), but not stigmatised since it is used by the most educated writers. It is essentially a socially acceptable, albeit imperfect (especially phonetically) representation of L which shows a high degree of uniformity and conventionality, thus, conforms to the requirements of a standard language.
• The concept of Cape Dutch, which was advocated by writers and academics such as Brink, Viljoen, Mansvelt and D'Arbez as an alternative to 'Afrikaans' as propagated by the GRA, was not entirely a language political fiction but was based on some (however extremely variable) reality, and it is thus not surprising that in the popular linguistic taxonomies a category described as 'Cape Dutch' or *beskaafd Afrikaans* existed. To characterise such variable linguistic usage simply as a random and ungrammatical mixture of Afrikaans and Dutch fails to recognise the structured nature of these intermediate forms.10

• Cognitive psychologists investigating the principles underlying the formation of categories, have shown that human categorising usually proceeds from a prototype or core which represents the clearest case(s) of the category. The category-core is surrounded by non-prototype members which show more or less similarity to the prototype, and which are perceived as being more marginal members of the category (Rosch 1973: 112, for further discussion see Final Remarks). Rosch & Mervis (1975) have argued that proximity scaling is an important tool for the study of categories: closely located items can be interpreted as representing prototypical category members, with prototypicality diminishing the further one moves from the core (see also Kempton 1978: 53 who argued that sharply bound categories characterised by clear prototypes appear as tight clusters in the graphs). Considering the multidimensional scaling plot given in figure 8.3, it is evident that clusters I and III, which have been described as relatively 'focused', show item-proximity. This facilitated their categorising in folk taxonomies (see § 3.3). Cluster II, on the other hand, is loose and thus not easily amenable to categorising, which possibly explains its absence from many popular classification schemes, as well as its occasionally pejorative classification as an ambiguous *tussending*, i.e. a category without clear prototypes.

8.1.4 Standardisation and Diglossia Revisited

The low level of within-cluster variance noticeable in cluster III (the emerging Afrikaans standard) indicates that around 1900 (i.e. before the onset of academy organised language

---

10 See, for example, De Waal's pejorative comments (1/11/1905 in *De Goede Hoop*) on the linguistic usage of Brink. More recently Ponelis (1996: 134) described the intermediate forms which regularly occurred in the periodical of the APO as '[d]utchified Afrikaans or approximate Dutch [...] essentially the product of poorly educated Afrikaans-speakers struggling to write and speak formal Dutch'. To recognise the structured nature of these intermediate forms is not to say that borrowing (from the standard variety) played no role in the process, and growing exposure to the standard language in a gradually improving school system probably contributed to the emergence of linguistic patterns different from those described for 18th century CDV. However, since the general structure of (especially morphosyntactic) variation continues patterns described for earlier texts (see § 6.1.2), it is likely that we are dealing with a case of stable variation. Borrowing, on the other hand, is typically *ad hoc* and upsets the systematic structure of non-standard linguistic varieties (Bailey 1980: 245).
planning/standardisation) 'Afrikaans' had developed into a relatively focused written variety, while there is evidence that the spoken language remained rather more variable (see Malherbe 1917: 15). The social composition of cluster III suggests that the agents of this focusing process were predominantly found among the intelligentsia, who also played a central role in early Afrikaner nationalism and contributed to the vernacular writing tradition (see § 3.1 and § 3.2). Alongside this new code developed a linguistic ideology which identified any influence from Dutch as artificial and non-authentic (see § 3.3).

However, only those who had ample exposure to the Dutch standard (usually via the education system) were in a position to recognise (and thus avoid or enhance) 'Dutch' features in their linguistic usage (as they had a clear typological conception of what Dutch was). That it was the élite who first showed fluent usage of a standardised L free from the influence of H, has also been noted for Greece by Frangoudaki (1992: 370): 'definitely since the 1950s, fluent usage of an elaborate D [Dhimotiki, L variety] free of K [Katharevousa, H variety] mixtures placed speakers in the highest social ranking, identifying them as members of an intellectual élite'; and similarly Kazazis (1992) noted that 'hyperdemonicism' was most dominant in written styles, while the spoken language of the Greek élite contained considerably fewer demotic forms (on 'hypervulgarism', i.e. the excessive use of L forms, as typical for non-casual styles see also Janda & Auger 1992). That the production of 'pure' (i.e. non-Dutchified) Afrikaans was the result of conscious efforts by the writers, is also suggested by the increasing number of excuses for 'incorrect' Afrikaans in private letters from about 1910 (see, for example, the documents quoted in Truter 1997: 96-97) as well as contemporary linguistic commentary. Malherbe (1917: 44), for example, comments with regard to the adjective inflection:

Eers in die laaste jare na veel studie en bewuste strewe, kan 'n mens 'n groot verbetering opmerk.
Dit gaan hand aan hand met 'n toenemende hoogskatting van ons taal wat op syjn beurt met jaloerse nouwgessetheid 'n so suiwer moontlike gebruik eis. Die 'Brandwag'-artikels vergelyk vandag pragtig met dié van vroëër jaargange. Tog word nog so gedurig foute gemaak (my emphasis).

The norms of the new standard were diffused through organisational networks linked to contemporary national-cultural movements. Not only the GRA decided to use Afrikaans for their proceedings (see Nienaber 1974), but Viljoen (1891: 3-4) also reports that Afrikaans was used in meetings of the Debats- and Christelijke Jongenvereeniging and among students.\textsuperscript{11} De Waal

\textsuperscript{11} This is also indicated in a letter by J.W.F. Grosskopf (4/11/1908) describing the South African students in Edinburgh who used Afrikaans as a kind of in-group code: 'Dit is net Stellenbosch of di Kaap, met diselfde bleddie Textbook en eksamen susteen, en met onbeperkte billiart, whiskey en poeier. Hulle praat wel deur di bank Afrikaans daar. Maar, God bewaar ons, hoe? En ook net om van die Skotsmanne nie verstaan te word nie. ', in the same letter he terms this variety 'Edinburgse Afrikaner-"slang"'.

described in his autobiography in anecdotal style how changes in language use and choice were diffused through cultural organisations.

Ongeveer in 1890, onderwyl ek student aan die Normaalkollege was [...] het ek aangesluit by die debatsvereeniging wat daaraan verbond was, en gevind dat al die studente, waar hulle nie hul toespraak in Engels afstreek nie, dit in Hoog-Hollands probeer doen het. Hier het my ingebore opstandige gees hom laat geld. Alhoewel ik self maar in die algemeen 'n slaaf van die mode was, het dit my daarem as eenmaal te erg voorgekom om onder mede-Afrikaner-studente by voorkeur die taal van Holland te gebruik; so het ek sommer uit die staanspoor uit weggeval in Afrikaans, in sy ongekunstelde vorm, al sou ek my met minder moeite in Hoog-Hollands kon uitgedruk het as die meeste van die ander. 'n Paar virde het my effens skee aangekyk, 'n paar het smalend geglimlag en een het 'n beledigende aanmerking gemaak, maar ander het dit goedsmoeds geduld, en weldra het ek volgelinge gekry. (1932: 20-21) [my emphasis]

It was mentioned in § 4.3.2 that members of class 2 (the new professional class) were socially as well as geographically mobile, and their social networks were thus characterised by a great number of weak ties. The importance of weak ties for the diffusion of innovations has been emphasised by the Milroys (1985b, 1992) who have argued that while dense and multiplex networks function as norm-enforcement mechanisms, weak ties facilitate the emergence and diffusion of innovations (see Milroy 1980: 45-52 on the structural characteristics of networks). Linguistic innovations typically originate at the periphery of the society's network, and are diffused initially through weak ties. They become established in society through their adoption by central members of close-knit networks who can provide an influential model for others, but who nevertheless have some weak ties with outside-members through which they are exposed to innovations (the Milroys call this group 'early adopters'). It is notable that many of those who contributed to the early vernacular writing tradition were marginal members of the Cape speech community, thus De Lima, Boniface and Abu Bakr Effendi were first generation immigrants, H.W.A. Cooper was English-speaking as was Meurant (although he later grew up in an Afrikaner family; see § 3.1), and within the GRA individuals of non-colonial birth still played an important role (Pannevis, Hoogenhout; see also Schoeman 1998: 226-227). Although multiple weak ties are prominent amongst those grouped as members of Cluster III, one would hesitate to classify them as innovators since they grew up in a time when 'Afrikaans' was already established as a

---

12 I prefer to view Jim and Lesley Milroy's work on social networks as a 'hypothesis' since the empirical evidence presented in Milroy (1980) is highly problematic, not only with regard to the way the network variable was conceptualised but also regarding the statistical analysis. Murray (1993) remarks that thirteen of the 27 t-test values reported by Milroy as statistically significant were not significant, and my own re-analysis of the data also failed to reproduce the statistical results. However, since the general idea that patterns of informal contact 'somehow' affect behaviour is intuitively pleasing, failure to reproduce results should not lead to the dismissal of a generally plausible hypothesis. However, more empirical research is urgently needed.
relatively uniform linguistic model in the popular discourse; yet they seem to lack the strong ties typical of early adopters. In a symbolic sense, however, the network ties of the members of cluster III were 'strong', since most of them had powerful nationalist affiliations. Following Anderson's (1991) argument that nations are 'imagined communities' whose members (although not directly linked) have a strong feeling of belonging which is created and reinforced through nationalism, nationalism can be said to transforms the weak ties which characterise modern societies (and which impede the establishment of new norms) into symbolically strong ties, thus facilitating the diffusion of new norms through the community.

As discussed in § 4.1.2 many of the members of cluster III used 'acrolectal' Afrikaans and 'acrolectal' Dutch in their correspondence in alternation, and the clear-cut distinction they maintain between the two codes (i.e. their texts can easily be classified as being intended as either Afrikaans or Dutch) suggests something one could term 'diglossic competence'. When Karl Krumbacher used the term diglossie in his book Das problem der modernen griechischen schriftsprache (1902) he supplemented it with many descriptive synonyms, such as doppelgesicht, doppelköpfigkeit and doppelsprachigkeit, which seem to express more clearly the individual correlate of societal diglossia, i.e. the co-existence of two grammars within the individual (or alternatively the existence of one grammar and a set of transformation rules; on the possibility of transformation rules in diglossia see Wexler 1971). In figure 8.15 the relative position of the different letter-corpora is indicated for two bidialectal writers: F.S. Malan (A) and Wynanda Hoogenhout (B): [1] Dutch and [2] Afrikaans.

Figure 8.16 Multidimensional scaling (see figure 8.3). Position of the bi-dialectal speakers F.S Malan (A[1], A[2]) and Wynanda Hoogenhout (B[1], B[2]).
The existence of such diglossic or bidialectal speakers within the corpus brings us back to the question of diglossia, which was approached in § 3.4 from a societal perspective, and it was argued that although diglossic tendencies existed and found expression in folk linguistic taxonomies, the general situation is probably best described as standard-with-dialects (or what Britto has termed user-oriented diglossia).

While the functional aspects of diglossia (i.e. domain complementarity) have been prominent in most studies of diglossic societies, Ferguson's original paper also contains several remarks on the structural (linguistic) differences which characterise diglossia. An important structural-linguistic criterion which is said to define diglossia is structural discontinuity, that is, the existence of an 'extensive' linguistic distance between H and L (and possibly M, see § 3.4). However, it was soon argued that even in the defining cases of diglossia structural continuity between the systems appeared to be norm rather than the exception. With regard to Greek diglossia Pappageotes & Macris (1964: 58) made direct reference to the possibility of a continuum situation, when they stated: 'The distinction between the Demotic and the Puristic is not absolute except at the two ends of the scale'. And Daltos (1980: 84, also Alexiou 1982) emphasised that while H and L exist as social constructs in the mind of the speakers, the linguistic analysis of actual linguistic data shows H and L to be idealised codes which are imposed on a linguistic continuum in which stylistic and social constraints interact. Discreteness, however, was not taken to be a requirement of diglossia by DeSilva (1974) when describing Sinhala diglossia as a continuum, and more recently Schiffman (1993/1997) defined Tamil diglossia as an example of what he calls 'heterogeneous diglossia'.

Between the styles [H and L] there are shadings from one style into another, i.e. it is possible to write modern literary Tamil with an archaic lexicon, but with nonarchaic grammar; it is also possible to swing in the other direction and to make modern literary Tamil more spoken, or to make educated colloquial more literary in flavour, or more nonstandardly colloquial. In any event, though linguistic cultures think of diglossia as either-or, it is often a gradient cline, with one variant shading into another. (Schiffman 1997: 210)

---

13 Ferguson (1959: 333); also Wexler (1971: 336), Fishman (1980: 4), Fasold (1984: 40), Winford (1985: 349, 351), Werlen (1988: 96). Britto (1986: 10-11) defined the structural distance between H and L as 'optimal', and opposed it to the structural difference between languages which would be 'super-optimal', and styles, which would be 'sub-optimal'. However, he admits that in practice the empirical decision of what constitutes 'optimal' distance is impossible (ibid.: 19-20, see also § 3.4).

14 The nature of the Greek continuum is, however, qualitatively different from the one discussed for the Netherlandic speech community at the Cape. In Greek inflected verbs, for example, cannot be assigned to either H or L since there exist forms which are combinations of H and L morphemes. Consider the following example for 'he was present' (given in Hawkins 1983: 226):

\[ \text{parusiástike (L form)} \rightarrow \text{eparusiástike} \rightarrow \text{eparusiásthike} \rightarrow \text{eparusiásthí (H form)} \]
In other words, while speakers (and linguistic cultures, i.e. speech communities) typically
describe linguistic variation either in bi-systemic or tri-systemic terms (i.e. acknowledging the
possibility of a category of well-defined 'intermediacy', see § 3.3 and § 3.4), the situation is
different 'in the field', where the linguistic systems employed by different speakers can often only
be loosely grouped, and individual lects shade into each other.

While it is today acknowledged that diglossia is more heterogenous than originally
assumed, the question remains whether its overall structure is by definition bi-systemic and
discrete, or whether a continuum of varieties can or should be classified as a case of diglossia.
The inclusion of linguistic continua into the diglossic model was rejected by Britto (1986: 14 ff.,
see also §3.4) who described H and L as 'diasesys' which can include any number of varieties
but are nevertheless structurally discontinuous (as well as functionally dichotomous). Following
Britto's interpretation of Ferguson’s (1959) paper, diglossia requires the existence of such
discontinuous diasesys (see figure 8.17), and a 'true' continuum (figure 8.18) bridging the two
poles of H and L (as suggested by Schiffman 1996) should not be classified as an example of
diglossia (see also Paolillo 1997).

Figure 8.17 Discontinuity: H and L as (variable) diasesys.

Figure 8.18 Continuity: Diglossia as a continuum of overlapping linguistic systems.

That the description of linguistic variability in diglossic situations has often been
impressionistic, based either on folk taxonomies or unspecific linguistic intuitions, rather than
empirical investigation, was emphasised recently by Paolillo (1997: 271).

While the different models of diglossia disagree as to whether they regard H and L as distinct, no
one appears to have developed an empirical procedure to test for distinctness in diglossia, or
especially in the urban centres of Cape Town, Kimberly and Port Elizabeth where the majority of immigrants settled (Watts 1976:46). From the 1860s social life in the colony was significantly anglicised (Bickford-Smith 1997). The historical records show that initially the Dutch-speaking colonists were more than willing to send their children to English-medium government schools, and it appears that until the 1850s/60s there was a consensus that English would be the future language of the colony, although the first dissenting voices were heard in the periodical press (see the source material presented in Du Toit 1985: 210ff.; also Steyn 1980: 124-131, Zietsman 1992: 4-6, Scholtz 1964: 27-28).

Dutch/Afrikaans-English bilingualism was most widespread amongst those who had received formal education (Colquhoun 1906: 144-145, Schonken 1914: 193), and frequently such bilingualism was transitional, a first step in a language shift towards English. Although there are not statistics for language shift towards English, we have anecdotal evidence of intergenerational language shift in traditionally Dutch/Afrikaans speaking families, as indicated, for example in a letter by E.J. Malan to her son F.S. Malan.

Oom Gawie (dokter) was verleden week hier by ons. Hy was 14 dagen te Wellington, zyn klein zoontje was zamen met hem, hy spreek altes Engelsch. (9/5/1893, see also letter by F.S. Malan to Viljoen, transcript in diary 4/11/1894).

At the turn of the century members of the Netherlandic speech community frequently used English as their language for (private) correspondence. It seems that English letters were typically written to siblings and friends, while letters to parents were usually written in Dutch, rarely in Afrikaans (see the discussion of the Louw family in § 4.1.1; Platt 1977 described a very similar pattern for Singapore and Malaysia, i.e. L1 with mother, L2 (English) with siblings and friends).

The growing bilingualism of members of the Netherlandic speech community gave rise to strategies of code-switching between Dutch/Afrikaans and English, which were facilitated by the fact that the surface structure of the two languages is relatively similar (at least in main clauses without auxiliaries; on the importance of surface structure equivalence for code-switching see Poplack 1980). The term code-switching (CS) is used here in the general sense in which Gumperz (1982: 59) has defined it, i.e. 'the juxtaposition within the same speech exchange of passages of speech belonging to different grammatical systems or subsystems'. In an early article Blom & Gumperz (1972: 424-426) introduced a distinction between situational and

---

17 The English-speaking part of the population was numerically in the minority during the 19th century, not reaching parity before the 20th century (Watts 1976: 43-44). While the population growth among the Dutch sector of the community was the result of natural increase (i.e. a higher birth rate than death rate), the English sector increased primarily through immigration.
conversational (or metaphorical) CS. Situational CS is akin to style-shifting in that it results from a change of the context in which the speech event occurs (generally change of setting or participants). Conversational CS, on the other hand, is not triggered by a change in the social parameters of the situation, but often relates to the discussion of particular topics or subject matters, though it can also occur without direct contextual motivation (Gumperz 1982: 60ff.). Myers-Scotton (1986: 406) termed this second type of CS 'switching as an overall unmarked [i.e. socially expected-A.D.] choice', a type of linguistic behaviour typical for many bi- or multilingual societies.

When more than one social identity is salient for the rights and obligations balance which speakers wish to have in effect for the current speech event, and each identity is encoded in the particular speech community by a different linguistic variety, then those two or more codes constitute the unmarked choice. [...] In this case, each switch in itself has no special social significance; rather, it is the overall pattern of using two varieties which carries social meaning. (1986: 406)

In other words, since the social identities of the speakers are associated with both languages, an overall pattern of CS is used to signal these bilingual identities. Myers-Scotton's model of the social motivation of CS is based on the structuralist notion of 'markedness'. That is, within a given community certain codes are 'unmarked' (i.e. expected, statistically normal) in certain situations. Using the unmarked code in a given situation X, implies a tacit acceptance of the rights and obligations holding between the participants, the choice of the marked (i.e. socially unexpected) code in such a situation is seen as an attempt to re-negotiate interpersonal relationships and one's group membership.

Although CS research has focused on contemporary immigrant communities and post-colonial societies, CS has also been described (among others) for Middle and Early Modern English (Schendl 1996, Wright 1998) and 17th-century German (Von Polenz 1991: 225-226). Mesthrie (1993) has argued that there is ample evidence (traveller's journals, poems, plays, descriptions of the education system as well as reported linguistic usage) indicating that code switching between Dutch/Afrikaans and English was a sociolinguistic reality at the Cape during the 19th century. However, his suggestion (1993: 58-59) that such a 'mixed code' was used habitually primarily by the Coloured inhabitants of the colony seems ill-motivated when considering not only the linguistic corpus of this study, but also numerous contemporary descriptions and literary sources reporting CS behaviour for the White section of the Netherlandic speech community (see below). 18

---

18 It seems, however, that in the 1930s when Afrikaner nationalism (and with it linguistic purism) dominated the scene, English-Afrikaans CS became increasingly associated with Coloured speakers, although it continued to be practised by Afrikaners (see Barnouw 1934: 29, 39-40). Today Afrikaans-
Both inter-sentential (at a clause or sentence boundary) and intra-sentential code-switching (within clause or sentence boundaries, including mixing within word boundaries) occur in the corpus. Much CS research makes use of the concept of a base or matrix language (ML) which is seen as providing the morphosyntactic frame for the code-switching utterances, as well as function words and inflectional material. The matrix language is usually identified 'on the basis of relative frequency of morphemes' (for a discussion of the 'ML frame model' see Myers-Scotton 1993a: 486-487, 1993b: 46ff.). Although it is not always possible to identify the matrix language of a given utterance unambiguously (see Giacalone Ramat 1995: 55, Gardner-Chlores 1995: 71, Muysken 1995: 181-182), the general idea of a linguistic matrix or frame has proved useful for the understanding of particularly intra-sentential CS. Most of the examples in the data show switches involving single English lexemes which are integrated into the morphological and syntactic structure of the ML (Dutch/Afrikaans). Myers-Scotton (1993a: 486, 1993b: 77-78) termed this type of CS 'ML [matrix language] + EL [embedded language] constituents', while Sankoff, Poplack & Vanniarajan (1990) referred to such one-off single lexeme constituents as 'nonce borrowings', Muysken (1997: 361) as 'insertion', Clyne (1987: 740) as 'transference' and McCormick (1995: 194) as 'code mixing'. Although there was a tendency in earlier work on code-switching to treat single EL constituents as borrowings and to exclude them from what counts as 'true' CS, most linguists now argue that single item CS and borrowing are not separate phenomena, but reflect the same process.¹⁹

Single lexeme switches in the data involve the following categories: NPs containing single nouns (1), nouns within an NP (2), adjectives within an NP (3), adjectives in predicate function (4), finite verbs (5), non-finite verbs in infinitival subordinate clauses (6), infinitives as part of a verbal cluster (7) (code switched items are italicised).

(1) ik het een van myn menschen afgevaardig om na een Property te kyk (Cillie, 1900)
(2) zy is een zeer knappe nurse (Katie Beyers 25/2/1906)
(3) want 3rd [class] rijdt elke common mensch (Rudolf Baalie 23/2/1894)
(4) ons is vandaag oor jollie want het regen nu (P.A. Goosen, sen. 27/7/1909)
(5) denk toch niet dat ik dit uit een boek copy (Sophie Hendrikz, circa 1900)

English CS is widespread throughout the Afrikaans speech community, yet appears to be particularly salient in the Coloured communities (see Carstens 1991: 274-275, Poonis 1993: 115; on code-switching in the Coloured community of District Six in the 1980s, see McCormick 1989, 1995).

dat ek jou hond aan Berth overhandig het om saam met sijn hond te train. (F.P.Hoogenhout 23/3/1912)

(7) More moet ik weer gaan 'call' en oermorgen moet ik naar een 'wedding' (Hester Hoogenhout 13/2/1906)

The use of inverted commas to mark lexical items is common in the corpus (see also Langenhoven 1929 [1922]: 49-50 on the generality of this practice), but not limited to foreign language material (see, for example, P.J. LeRoux, circa 1907 Vader zyn 'pensionen geld'). Turning to the social motivations for these single item switches, many of them relate to specific topics, such as transport technology (8), employment (9), social life and entertainment (10) and education/school environment (11). CS is also used to reiterate information (12).

(8) [D]e vrouw ging met de kleinere kinders in een carriage (Rudolf Baalie 23/2/1894)
Ons het gegaan met Mr Brink z'n motor bike en side-car (W. Hoogenhout 15/6/1914)

(9) zoo dat ik kan weet om mine notice te geven (Sophia Daniels 23/9/1900)

(10) Hy het gestuur 'n pragtige 1/2 doz china tea cups 6 silver lepeljies en 'n prachtige stand met teapot en di spirit lampie. (Sebastina Goosen 19/4/1910)
Ik heb Susie 'Chess' leeren spelen en nu zal zy my 'painting lessons' geven. (Johanna Hoogenhout 26/10/1902)

(11) Mr de Kock zeg dat Jacob en Jannie gepassed hebben. (Susie Goosen 23/2/1909)

(12) Waar is logies om al de mensen met de ordinary of gewone treinen [?] (C.J.Van der Merwe 30/11/1900)

Example (11) shows CS within word boundaries, i.e. the lexical item comes from the EL but the inflectional material from the matrix language. Such 'mixed constituents' are quite common in the corpus, and occur mostly in the past participle (see also Te Winkel's 1896: 361, commentary on CS behaviour where he gives the following examples: ons het ons baing geënjoy and ons het gewalk).

(13) ik heb [...] nog een weinig by hem georderd (J.J. De Villiers 4/12/1882)
dat vir hulle almal spesiale eerste Klas sitplekke ge-'book' is (J.W.F. Grosskopf 21/10/1906)
Ek het vir hom en Ruth gezien en by hulle gesupper in 'n restaurant. (F.S.Malan 2/4/1896)

CS in the corpus also involves longer constituents, such as EL idioms (14), as well as longer strings of EL material following EL morphosyntactic rules ('EL islands' in Myers Scotton's terminology), see example (15).
Although there are many examples of single item code-switching and some examples of EL islands in the corpus, in terms of relative frequency the occurrence of code-switched material is rather rare, generally limited to one or two switches per text. There are, however, some exceptions, most notably Johanna Brümmer, the young fiancée of F.S. Malan. In one letter in particular (10/5/1895) she regularly inserts English EL constituents into the Afrikaans frame. The following is a characteristic extract from the letter.

Example (16) corresponds to Myers-Scotton's (1993: 482) discussion of CS as an unmarked choice, i.e. it occurs in a peer-group situation and is characterised by frequent intrasentential switches. Intersentential switches are also common in Johanna Brümmer's letters where they often occur in a kind of topic-comment structure. In example (17) the main topic is discussed in Afrikaans, followed by a metalinguistic commentary in English, then the switch to English is commented on in Afrikaans. The rest of the letter is entirely in Afrikaans (on CS in topic-comment constructions see Romaine 1995: 163).

Use of intersentential CS in topic-comment constructions occurs also in F.S. Malan's letters to Johanna.

Topic-comment CS occurs also in the letters by Mimmie Laubscher, a teacher at the prestigious Bloemhof school in Stellenbosch, to her boyfriend Daniel Marais Hoogenhout. In example (19) and (20), the discourse topic is described in English, and Afrikaans occurs in the comment position.
P.S. By the way I have not had a photo taken yet - *ik is zoo bang di glas zal bars* - but I hope to have one taken soon, and will then let you have a copy. (22/5/1909)

If you gave me full permission to kiss him do you think I would? *Dit zal erger als een 'seep-schottel' smaak, na die leker laste zoen van jou bij die trein die aand.* (9/8/1909)

In some cases the postscript following a Dutch or Afrikaans letter is written in English (Maria Malan 14/5/1892), and reference, for example, to an English book or letter which was received can trigger a complete switch to English half way through a Dutch letter (Rijkie Louw 4/8/1909).

There are some literary examples of English-Dutch/Afrikaans CS for the 19th and early 20th centuries, such as the popular poem/play *Kaatje Kekkelbek* written by A.G. Bain in 1838 (see Mesthrie 1993 for a discussion), Brink's play *Volk en Taal* (1905) and the short-stories by Lub (1908). While in *Kaatje Kekkelbek* the protagonist is the stereotypical 'Hotnotsmeid', later representations of CS (such as Brink and Lub) identify young Afrikaner women (and to a lesser degree young men) as those typically engaged in CS behaviour (see also Viljoen 1891: 4, De Vos 1891: 22-23, Te Winkel 1896: 371). Ridiculing and criticising, what Mansvelt (*Zuid Afrikaan* 1/5/1890) termed *Mixpickles Afrikaans*, was a common topic in the metalinguistic discourse at the end of the 19th century, as illustrated by the following extract from an anonymous article in *Ons Land* (25/10/1895).

> Hoewel gij al dadelijk moet genoicic hebben, dat ik geen advocate ben van zulk een language, wil ik nu ronduit zeggen, dat my object is om allen, de guilty moet plead aan 't gebruik van *Kaapsch*, soo thoroughly disgusted met zichzelven te maak, dat hulle in die future liever hulle mond zal hou, als hulle dan nie een fatsoenlijker language kan employ.*

The evidence from the corpus gives some support to the 19th century stereotype that English-Dutch/Afrikaans CS was prominently employed by the younger generation (age group 2). There is a tendency in the corpus for CS to be prominent in the letters written by young women, Johanna Brümmer and Mimmie Laubscher being principal examples; however, it is not absent from letters written by young men (such as F.S. Malan, Gideon Goosen and J.F.W. Grosskopf). CS between Dutch/Afrikaans and English does, however, not occur in letters written by members of the older generation (age group 1), with one exception only (Rudolph Baalie, lay preacher of the Moravian mission station in Genadendal).

Bilingualism not only leads to the *ad hoc* insertion of lexical material from the EL into...

---

20 Returning in passing to the question of folk taxonomies and labelling, it is conspicuous that the term *Kaapsch* is used here to refer to the mixed code itself. Similarly McCormick (1989, 1995) has shown that today the term *Kombuistaal* is used by the community of District Six to cover 'both non-standard Afrikaans, and switching between English and Afrikaans strings' (1995: 199).
the ML, but also to syntactic interference (see Haugen 1956 for a discussion of the differences between CS and interference, the latter understood as the ‘overlapping’ - and not the alternation-of two languages). So-called ‘anglicisms’ were common in the 19th and early 20th century, and reportedly pervaded many early Afrikaans literary texts (see Pienaar 1931: 173-174, Nienaber 1936: 141). Although this is not the place to deal with such contact phenomena in detail, some comments should be permitted, especially as this area has largely been neglected in diachronic (and synchronic) studies of Afrikaans (Donaldson 1991: 7). As already argued in § 2.1, by placing the formation of Afrikaans into the late 18th century, the South African Philological School has interpreted the 19th century as a time of consolidation and stasis. This position has been challenged by Donaldson (1995: 223, see also Valkhoff 1966: 195), who suggested that with respect to language contact phenomena ‘the linguistic transformation that would take place after the British occupation of the Cape in 1795 was to be as great as, if not eventually greater than, all the changes that had taken place hitherto.’ Taking Swart & Prinsloo’s (1936: 1-70) and Donaldson’s (1991: 170-177) lists of common loan translations as a point of reference, several of these were found to occur in the corpus, indicating that many of these loan translations are possibly not so much ‘the result of spontaneous translation by the bilingual Afrikaner’ (Donaldson 1991: 170) but have a certain historical depth.

(21) so ver (Engl. so far; up till now)
met de boerderij gaat het nog zo ver alles goed (E. Smith, circa 1900)

(22) heel dag (Engl. all day)
en natuurlik ons zien hom heel dag (C.Beyers, circa 1918)
i k loop maar heel dag in de tuin (B.J. Brümmer 19/9/1907)
want ik word heel dag daarna gevraagd (C.J. Van der Merwe 30/9/1900)

(23) skaam voel (Engl. to feel ashamed)
i k voel altiid zo skaam fir my briewe (Hester Hoogenhout 13/2/1906)

(24) een van die dae (Engl. one of these days)
een van die dage gaan ons Dr. Newnham vra (Johanna Brümmer 10/5/1895)
een van die dagen kom ik weer thuis (Wynanda Hoogenhout 21/9/1905)

(25) lyn (Engl. line)
net een paar lynen om u te vragen (P.J. Le Roux 18/2/1908)
in haas net gou 'n paar lijntjies (Hendrina Hoogenhout 13/5/1919)

(26) verlang vir (Engl. to long for)
en kan part dae vreselik vir die plaas verlang (R.I.Steyn 4/12/1919)

(27) lyk (Engl. to like)
as u nie lyk dat ik van hier een order moet stuur (Engela Furter 5/10/1909)
Ina sal dit baing lyk hier (Wynanda Hoogenhout 7/12/1908)
(28) **bly** (Engl. to stay)
   vandat ik in die Pastorie blijft (Maria Cronje 30/7/1906)
   soals u weet bly ons nou in die Kaap (Susie Baartman 6/3/1920)

   Rather common is also the use of **meen(en)** (Engl. to mean) for Dutch/Afrikaans
   **bedoel(en)** and **beteken(en)** (Swart & Prinsloo 1936: 30, DeKlerk & Bosch 1998: 47), and the
   reflexive extension -**self** which is used in Dutch only in emphatic function, but has been
   generalised under English influence to cases where no emphasis is implied (Ponelis 1979: 83,

(29) hertelyke groete van Martha en **myself** aan u allen (E. Smith, circa 1900)
    Theron, Watermeyer en **jouself** ni te vergeet ni (N.J. De Wet, June 1895)

   In her discussion of language use in District Six, McCormick (1995: 205) has commented
   on examples of loss of verb final from constructions where the VP consists of a modal or tense/
   aspect auxiliary + main verb (such as: **Ons moet study altwee vs. ons moet altwee bestudeer**), a
   tendency which has been attributed to influence from English (see also Donaldson 1991: 276).
   Examples of this occur in the corpus but are rare, and it is not always clear whether the
   construction is not a result of the possibility of **Ausklammerung** or extraposition, i.e. the
   movement of heavy material (particularly PPs) to the right of the verb, which has been described
   for dialectal Dutch and Middle Dutch (see Lebbe 1997). In the corpus there are isolated examples
   of main clause word order (V2) after the complementizer **dat**. Donaldson (1991: 276-277)
   interpreted the omission of SVO in subordinate clauses to influence from English, and examples
   such as **Mama werk omdat sy is 'n onderwyser** clearly show interference from English. However,
   the examples of V2 after **dat** which were found in the corpus show VPs consisting of an auxiliary
   +main verb with the main verb in sentence final position, a structure which cannot be attributed
   to influence from English (thus, for (30) English influence would probably have led to
   constructions such as **zoo dat ik kan doen nachtwerk**, i.e. so that **I can do nightwork**).

(30) **zoo dat ik kan nachtwerk kan doen** (L. Bosman 2/1/1900)
    laat dit zal voor de gemeente gedrag wordt (C.Coopman 18/9/1909)

   The construction used by Louisa Bosman is particularly interesting as it shows a tendency
   described by Wafer (1994a, see § 7.3.1) for 18th century CDV, namely the movement of
   auxiliaries (as well as adverbials and negation particles) to the left, while occurring
   simultaneously in the original position.
In sum, English influence is considerable in the corpus, which supports Donaldson's claim that in terms of language history, the sociopolitical dominance of English during the 19th century is important for an adequate understanding of language history at the Cape. From the 1830s children were exposed to English in the public school system, and it has been emphasised repeatedly that with the rise of literacy levels from the late 19th century reading material available to the literate population was predominately in English, thus deeply entrenching English among the educated classes.\footnote{Calquhoun (1906: 147), Barnouw (1934: 38), Langenhoven (1938 [1914]: 33), Varley (1952: 20); see also the contents of the private library of the NGK minister David Malan (husband of Katie van Huyssteen), SAL-MSB 852-1-23.}

8.2.2 Patterns of Code-Choice and Code-Switching: Dutch- Afrikaans

It was noted above (§ 8.1.4) that some of the individuals sampled for the corpus are best described as bidialectal speakers, that is, for them Dutch and Afrikaans appear to have been clearly separated social and linguistic constructs, and their texts (or stretches of discourse within their texts) can be assigned relatively unambiguously to one system or the other. While the first part of this section will focus on the overall pattern of code-choices of some of these bi-dialectal speakers, an attempt will be made in the second part of this section to account for the structure of variation in the data from a CS perspective, drawing on Muysken's (1995, 1997) recent discussion of CS processes. There is some debate as to whether switching between dialects should or can be analysed from a CS perspective, and the difficulty in identifying the actual point of switching between closely related dialects or languages which share many lexical items, makes dialectal CS analysis difficult. However, while Labov (1971: 461-462, 1991[1972]: 189-190) questioned the usefulness of CS analysis for such data, others (Blom & Gumperz 1972, Giesbers 1989: 4, 23-24, DeBose 1992, Romaine 1995: 170f., Giacalone Ramat 1995) maintain that a CS approach to dialectal data is not only possible, but also relevant for both our understanding of dialectal variation as well as for our understanding of CS in general.

Of particular interest with respect to the patterns of language choice is the correspondence between F.S. Malan and Johanna Brümmer, as well as the family correspondence of the Hoogenhout family. F.S. Malan's diaries are an intriguing sociolinguistic document as he uses English, Dutch and Afrikaans in strict alternation, i.e. if one entry is written English the next will be written in either Dutch or Afrikaans, the next one again in English and so forth. Initially (diary 1892) the alternation is only between Dutch and English, but in 1893 when he was studying in Edinburgh, Afrikaans joined the repertoire and became slightly dominant (in terms of
frequency). From 1895, however, code alternation was again limited to Dutch and English, with English gaining ground. The letters to his fiancée Johanna Brümmer show a similar pattern, that is, increasing use of Afrikaans from 1893, peaking in 1894; during the same time period he wrote few letters in Dutch and English. From 1895 his use of Dutch and English increased again, but English disappeared entirely during the Anglo-Boer War, probably for political/ideological reasons. After 1897 there are no longer entire letters written in Afrikaans; however, occasionally Malan code-switches for one or two sentences into Afrikaans, typically at the beginning or end of letters (see, for example, the letter dated 27/12/1906).

![Graph showing code choices in F.S. Malan's letters to Johanna Brümmer (N = 361 letters) 1891-1920 (in percentages, E=English, A=Afrikaans, D=Dutch).](image)

During his 'Afrikaans period' (1893-1895), he repeatedly urged his fiancée Johanna Brümmer to write to him in Afrikaans. Johanna, however, shows a clear preference for English, and when she started writing in Afrikaans it was to please and appease him (see her letter dated 23/11/1894). The compromise suggested by Johanna is similar to the code-alternation practised by Malan in his diary.

You don't mind my writing English this week [?] I hope always to write Afrikaans every other week. Are you satisfied with this arrangement? I hope so. (8/3/1895)

---

22 That South African students frequently developed an interest in Afrikaans when studying in Europe was suggested by J.F.W. Grosskopf (see letters dated 4/11/1908, 18/2/1909).

23 During the Anglo-Boer War F.S. Malan was imprisoned with hard labour for the publication of a letter describing the conditions in the Howick concentration camp in the newspaper *Ons Land*. The imprisonment prevented him from taking up his parliamentary seat for the *Afrikaner Bond* (see § 3.2).
Yet, judging from the quantitative overview of code-choices by Johanna Brümmer, this agreement was on paper only, and did little to minimise the dominance of English in her repertoire. In analogy to the development of F.S. Malan's repertoire, English disappeared as a code-choice after the Anglo-Boer War.

![Figure 8.20 Code choices in Johanna Brümmer's letters to F.S. Malan (total=209 letters) 1893-1904 (in percentages, E=English, A=Afrikaans, D=Dutch).](image)

English as a code choice is absent from the Hoogenhout family correspondence. C.P. Hoogenhout was one of the founding members of the GRA (see § 3.2), and the correspondence of his children was conducted predominantly in Dutch and Afrikaans. Interestingly Hoogenhout's sons use Afrikaans only in letters to friends and siblings, but never in letters to their parents who they always address in Dutch. At the same time their sisters Johanna and Wynanda, as well as later their wives, frequently use Afrikaans in letters to C.P. Hoogenhout and his wife Catharina Maria. Hester van der Bijl, for example, who married Nikolaas Marais Hoogenhout, introduced herself to her future parents-in-law using Afrikaans (while the other daughters-in-law started off in Dutch), and Afrikaans remained her dominant language for family correspondence. However, the fact that she repeatedly justified and explained her code choice suggests that in this case the use of Afrikaans was (at least initially) what Myers-Scotton has termed a 'marked choice' motivated by a desire to minimise social distance and to express a feeling of familiarity.

Weet u wat. Ek virlang regtag om U allen te zien. Is dit ni wonderlik van mij[?] Daarom schrijf ek ok sort van Afrikaans, omdat dit fir mij is, net of ek u al lang ken. (circa 1905, see also letter dated 13/2/1906)

Her children, on the other hand, wrote to their grandparents in Dutch as late as 1920 (see
letters by Pieter Hoogenhout 5/12/1920, and Klaas Hoogenhout 5/12/1920), and she overtly discouraged their use of Afrikaans (see letter 10/11/1906).

There are not many letters available for Johanna Hoogenhout, who died in 1905, but for the younger daughter Wynanda there is sizable collection of letters to her parents. Wynanda Hoogenhout's code choices show a clear pattern towards increasing use of Afrikaans from about 1910; possibly motivated by a kind of linguistic insecurity since her Dutch letters (1905-1909) contain several grammatical mistakes which were corrected in pencil by her father (and possibly remarked on in his replies). The graph given in figure 8.21 indicates that the choice between Dutch and Afrikaans was for her at least at times complementary, i.e. in the initial years when she used primarily Dutch she wrote few Afrikaans letters, and vice versa. Only during the years between 1916 and 1918 did she use both codes more or less equally, and those years appear to mark a transition period before her complete shift towards Afrikaans.

Figure 8.21 Code choices in Wynanda Hoogenhout's letters to her parents (total= 155 letters) 1905-1921 (in percentages, D=Dutch, A=Afrikaans).

Figures 8.19 - 8.21 illustrate the possibility of different code choices in cases where the context (degree of formality and interlocutor) was stable, i.e. code choices were not situationally motivated but served to negotiate and establish identities and extra-linguistic alliances in the historical context of Afrikaner nationalism (on linguistic 'acts of identity' see the classic study by LePage & Tabouret-Keller 1985, for a more recent discussion Johnstone 1999, Johnstone & Bean 1997). However, code alternation could also depend on situational factors. In the Hoogenhout correspondence, for example, code choice was dependent on interlocutor for the male children, i.e. Dutch to parents and Afrikaans to siblings and friends. Use of Afrikaans also appears to have been a common choice for correspondence between Afrikaner nationalists:
Ik hoop jy neem my nie kwalitee dat ik jou in die Patriot taal, ons taal geschreywe het nie, ik zien dat jy en Hofmeyr zoo correspondeer, en ik schryf ook graag zoo, gewoonlik schryf ik aan Bondmannen in die taal - (D.C. De Wet 24/11/1888)

That degree of formality also affected code choices was shown by Roberge (1994a) in his comparison of Louis Trigardt’s diary with an ‘official’ letter written by Trigardt, in which he employed a language considerably closer to standard Dutch. As argued in § 4.1.2 the stylistic dimension (understood following Labov as degree of formality) was kept constant in this study, and texts approaching the formal side of the stylistic continuum were not sampled systematically.

Code alternation between Dutch and Afrikaans not only occurred across different letters by the same writer, but also within letters. Typical is the use of Afrikaans in cases where direct speech is reported (‘quotational code-switching’, Romaine 1995: 175, see also Gumperz 1982: 75-77). This type of (inter-sentential) CS also characterises literary texts such as Benigna (1873, see § 3.1) and Lub’s short-stories (1908). In the corpus quotational CS is most common when representing the speech of children, as in the following extract from a letter by Johanna Brümmer.

(31) Miss de Villiers slaapt nog elken nacht by my. Jacques is net groot vrienden met haar. Als zy s’morgens weg gaat zeg hy altyd “Auntie moet weer kom.” Nu toe ik hem vroeg wat ik aan u moet schryven zeide hy: “Ze ver pappie ik leer my lees.” (14/5/1901, see also Johanna Brümmer 22/12/1905, Susanna Goosen 11/3/1918, letter by Janie van Huyssteen 31/12/1907)

CS into Afrikaans also occurs in letters (or parts of letters) addressed to children. The letters by Maria Malan to her sister-in-law Katie are all in Dutch; however, on three occasions she shifts to Afrikaans when providing information intended for Katie Malan’s children.

(32) Wat doen de kleinen? Wat zegt vrede van zyn briefje? ze ver hem en Rechtie hulle moet oupa kom help werke, arme oupa heeft zoo veel te doen. Nu liefste, myn tyd is beperkt, pa wacht om de brieven te posten. (29/5/1906, see also her letters dated 4/11/1906, 22/12/1907)

Narrating anecdotes can also trigger a switch into Afrikaans, as evident in a letter by F.S. Malan (10/4/1892, see also his diary entry 8/12/1892). It appears that this strategy was not limited to members of the Netherlandic speech community, but also employed by English-speaking South Africans.

It is often complained by persons lately from England, that when the English South African has a joke to make, or comic story to tell, he lapses into the Taal, which is not understood by the newcomer; the truth being that it is the use of the Taal which transforms an ordinary sentence into a joke, and makes the simplest story irresistibly comic. (Schreiner 1992 [1923]: 79)
A general pattern of Afrikaans-Dutch CS 'as an unmarked choice' is exemplified in some of the Dutch (ML) letters by F.S. Malan (10/4/1892, 1/6/1893) which show CS between Dutch, Afrikaans and English (CS into EL italicised [Afrikaans] or small caps [ENGLISH]).

(33) Kalfsoog is deur! Die results is uit! Ik weet niet of dit nou juist zoo danig aangenaam nieuws zal wezen, maar myn hart kan ook bly wezen. Of ik het verdiend heb weet ik niet maar het is toch daarom lekker. Ek wil Joe haar hart ook 'n beetje wakker scheer. Denk er aan, of Kalfsoog het ook 'n ou examentjie gemaak. Ik had gehoopt om in Class II te komen doch daar er geen Class I is komt het niet zoo groot op aan. [...] Pas 'n brief van Hendrik Theron van London ontvangen. Onder andere schryft die stonkerd: 'Arie, hoe better zal die kleiding ze harije klop in Afrika als sy dit hoor. Ja, ni, dan droom sy zomaar weer lät sy ver hom zoo in zyn Toga zien staan.' waar hy er aan komt weet ik niet. Het schynt my of hy meer weet dan ik. Wien kan hy bedoelen? Wiens hart zal om ou Kalfsoog klop? 'FISHING AGAIN' YOU WILL SAY. [rest of the letter in Dutch] (1/6/1893)

It is difficult to pinpoint individual switching sides (i.e. the actual point of switching) in such examples of dialectal CS as the extract contains many 'homophones morphemorphs' (Clyne 1987: 754), i.e. morphemes which are identical in both codes (such as is for the 3rd person singular as in 'Kalfsoog is deur!'), and the typographical marking of CS by italicisation can thus only be seen as a very general guideline.

Switches in example (33) involve both grammar and lexicon, are multi-constituent, and are typically located at clause boundaries. Muysken (1997: 361) has discussed this type of switching under the term 'alternation', which refers to a switching pattern characterised by switches involving complex structures (typically several constituents in a row). Alternation can be distinguished from 'insertion' (i.e. the embedding of EL constituents into a clearly defined ML) and 'congruent lexicalisation' (i.e. the continuous mixing of material from two languages, with no ML identifiable). Both 'insertion' and 'congruent lexicalisation' will be discussed in more detail below.\(^{24}\)

Example (33) also shows possible instances of triggering, i.e. cases where 'an item of ambiguous affiliation (that is, one that belongs to the speaker's two systems) triggers off a switch' (Clyne 1987: 744, 754-55, see also Myers-Scotton 1993: 491, Romaine 1995: 229-230, Giacalone Ramat 1995: 59). In this case the word 'lekker' belonging to both systems but having a wider range of application in Afrikaans, appears to trigger the next utterance in which the infinitive (following the modal wil) remains uninflected. Note that the construction Joe haar hart (rather than standard Afrikaans Joe se hart) is typical for F.S. Malan's use of the periphrastic possessive in his Afrikaans texts (while he avoids the periphrastic construction when writing Dutch). A

\(^{24}\) Note that Muysken describes 'alternation' as a type of intrasentential CS, while it is here understood in a more general sense as any switch which involves the switching of multiple constituents, whether intra- or intersentential.
second trigger is the use of an English expression (reported as direct speech) which triggers a switch to English, and the sentence is finished according to the morphosyntactic rules of English.

In those letters with Afrikaans as an ML (generally cluster III), alternation into more Dutch-like varieties is frequently related to topics which are connected to domains dominated by (standard) Dutch, especially education, and religion (34), while a switch towards Afrikaans can be triggered by topics connected with the private domain of the family (35); CS into EL italicised. Again switching occurs typically at clause boundaries or the periphery of sentences.

(34) Mij zendeling vriendinne en ek het net al zwaar gevoel over die tekort, nog gister onving ik een brief van haar, waarin zij mij aansporen om tog nog meer en meer te bidden voor die Werk. [letter continues in Afrikaans] (B. Smit 23/11/1917, see also Hester Hoogenhout 8/11/1905)

(35) [W]ij zijn gelukkig en tevreden. Winnie maak zulk lekker kos en is zoo handig in al hare huiswerkjes dat ik waarlijk verwonderd ben, en er is dus geen gevaar dat ik eenige neiging zou hebben om naar 't hotel terug te gaan. [letter continues in Dutch] (Andries Hauptfleisch 11/6/1914)

That alternation between Dutch and Afrikaans was not limited to the written register, but also occurred in spoken language interaction, is suggested by an anecdote narrated by Hester Hoogenhout in which her son Pieter (born 1908) is shown to have both codes at his disposal, alternating between the two in the course of the same situation.

Laatste Maandag was hier 'n klein Hollander meisje, en sij moest help 'n huisie van petrol blikke bouw, dit was in die middag, en was in die kamer, maar kon alles hoor deur die oop venster, Pieter was eigentlik die bouwer en sij en Klasie die draërs, dinge het miskien te stadig na sij smaak gegaan, toe hoor ek hem uitroep na die klein meisie, met 'n oprege Hollandse uitspraak, 'Antonia, breng twee binnen tegelijk dan is het wel spoedig al gedaan' dit was te mooi om te hoor, en gewoonlik praat hij sì industri隐蔽 haar. Op 'n ander keer was het weer 'heh! daar zit nog wat leuke sousjes in de pan.' (10/11/1916)

Muyssen (1997: 364) suggested that processes of 'alternation', i.e. relatively long multi-constituent switches, are 'particularly common in stable communities with a tradition of language separation', and Cochran (1997: 49-50) has argued that the existence of such alternation patterns can be seen as a diagnostic feature of (classical) diglossic communities.

While those individuals which were characterised as 'bidialectal' in § 4.1.2, show a CS pattern of 'alternation' (i.e. multi-constituent switches, typically at the periphery of the utterance or intersentential), others appear to randomly mix elements of the two systems, but never conform to any one system for a longer stretch of discourse. Muyssen (1997: 362) termed the latter type of switching 'congruent lexicalization', and described it as 'a situation where the two languages share a grammatical structure which can be filled lexically with elements from either language', and where the switching process is best characterised as a continuous going back and forth between the two codes, and the absence of any clearly defined ML. According to Muyssen
congruent lexicalization is 'akin to [...] monolingual variation' (ibid.), and can be described quantitatively using Labov's notion of the linguistic variable. The examples (36) and (37) do not in any way constitute a special case but are representative of the linguistic behaviour of several individuals in the sample.

(36) ik heef nu tog een brief van Edmund Smith uit Barkely gekregen [.] het gaan met hem en Martha reg goed ook met Johan en Barend gaat het reg goed (B.J. Brümmer 7/12/1901)

(37) zaterdag heeft ons bazaar [.] dan ver wacht wij veel menschen [.] zij hebben mos een plan om hier een ge meenten te stigt [.] maar of zij het zal regcht krijgen weet ik niet (Petronella van Huyssteen, circa 1906)

Congruent lexicalisation, as exemplified in (37) and (38), does not involve long complex structures and typically occurs within constituent boundaries, i.e. within the VP (ik heef gekregen; as discussed in §6.1.1 het and heeff(i) were treated as allomorphs for the purpose of the morphosyntactic analysis), as well as between NP+VP where the NP is the head determining INFL (wij verwacht, het gaan). CS within constituents is absent from the texts produced by bidialectal speakers, but dominates amongst the individuals grouped in cluster II (§ 8.1.1). Members of clusters I and III, on the other hand, not only engage in code 'alternation' involving switches between long stretches of discourse (examples of which were excluded from the quantitative analysis), but also in processes of 'insertion', i.e. occasional single item switches within a clearly identifiable and statistically dominant ML, which accounts for the (relatively minor) quantitative variability in their linguistic behaviour.25 The three-cluster solution, established on quantitative grounds using the construct of the linguistic variable, thus holds when approaching the variability in the data from a CS perspective since the three clusters relate to different processes or patterns of CS.

Cluster I  insertion and alternation
Cluster II  congruent lexicalisation
Cluster III insertion and alternation

Alternation and congruent lexicalisation differ not only structurally, but probably also with regard to their sociolinguistic and neurolinguistic embedding. Alternation, as discussed above, correlates with social stability and code complementarity, while congruent lexicalization is typical for 'second generation immigrant groups, dialect/standard and post-creole continua'.

25 Note that Muysken's distinction between 'alternation', 'insertion' and 'congruent lexicalisation' is not meant to suggest the existence of clear-cut categories (1997: 377), and the above discussion should, therefore, be seen as reporting trends, rather than postulating unambiguous patterns of CS.
Although there is still 'no consensus of opinion on the issue of neuroanatomical organization in bilinguals' (Romaine 1995: 91), some speculations on the psycholinguistic aspects should be permitted. Studies of bilingual aphasics support the hypothesis of the existence of two I-languages for bilinguals (see Paradis 1985: 22), and it is generally assumed that during CS (bilingual or bidialectal speech mode) both languages are activated (Grosjean 1989, 1995). Muysken (1997: 364) suggested that the different types of CS have possible psycholinguistic correlates with respect to the 'level of monitoring in both languages, the triggering of a particular language by specific items, and the degree of separateness of storage and access system'. Thus, during alternation activation probably switches from one language to the other, while during insertion the dominant language might be deactivated temporarily. Regarding congruent lexicalisation, however, it seems that the two (or more) languages or dialects do not constitute distinct codes, separated by a kind of mental barrier, but have 'merged', and one can speculate that the neurolinguistic distinction of two I-languages is somewhat (if not entirely) blurred in such cases.

8.3 Die Triomf van Afrikaans

In this chapter an attempt has been made to integrate the results from the multivariate quantitative analysis with observations from a CS perspective. This not only allowed me to comment at least briefly on some of the results of Dutch/Afrikaans-English bilingualism, but showed that the three cluster solution which was established on the basis of quantitative criteria also related to qualitatively different processes of CS. Interpreting the (marginal) variability in clusters I and III as the result of insertion strategies allows us to see the change from (standard) Dutch (cluster I) to (standard) Afrikaans (cluster III) in terms of a ML-'turnover' (Myers-Scotton 1993b: 70-74, 220-226), i.e. a change in the directionality of insertion processes. This change can be understood as the result of a language dominance shift in H domains, that is, from a predominantly Dutch system with 'Afrikaans' insertions (cluster I), to an Afrikaans system with 'Dutch' insertions (cluster III).

Historically there is evidence to locate the change towards wider use of Afrikaans into the years after 1910-1914 when Afrikaner nationalism gained large-scale support across social groups (see Hofmeyer 1987: 107, 110-111). At the same time, language political changes took place: in 1914 Afrikaans was introduced as MoI in primary schools, in 1915/1917 the first normative spelling rules were published by the Akademie and the publication of Afrikaans books (up till then relatively minor) took off (see appendix C). Those involved in the taalstryd had little doubt that this change was for good.
Wat moet dit uw hart tog bly maak om te sien, how wonderlik snell die beweging voor Afrikaans vorder in die laatste tyd. Dis byna ongelooflik, dit maak 'n mens byna bang, dat dit nou met een weer al te vinnig gaat. (Letter by W.M.R.Malherbe, editor (?) of Die Brandwag to C.P. Hoogenhout, 19/5/1916, MSC 1-37-2)

This interpretation is supported by the diffusion of Afrikaans forms across time in the sample. Thus, when grouping the sample texts into four successive time periods (1880-1890, 1891-1900, 1901-1910, 1911-1922), the percentage of Afrikaans forms (all variables combined) shows a steep increase only after 1910.

![Graph showing percentage of Afrikaans forms across time](image)

Figure 8.20 Percentage of Afrikaans forms across time (1880-1922).

The rise of Afrikaans as a new standard language has been described above as an example of a language dominance shift, which (like any other variety-shift) can be said to constitute a case of language spread (Ferguson 1988:120, 126). Considering that several individuals in the corpus had both (acrolectal) Dutch and (acrolectal) Afrikaans at their disposal, their shift towards Afrikaans can be described in terms of a (clear-cut) language turn-over. However, for those whose linguistic behaviour was more mesolectal (cluster II) linguistic insecurity continued, and as late as 1932 (when the Van Huyssteen children had replaced Dutch with Afrikaans as their language of correspondence, see the letters in MSC 853-5-1) Hendrik van Huyssteen's second wife Hester apologises for her inadequate knowledge of the proper standard (her letter conforms on the whole to the variable patterns described for cluster II) to her stepdaughter Katie: 'Lees tog maar al myn fouten recht hoor!' (6/8/1932, MSB 853-5-1).
EPILOGUE

A problem without a solution may interest the student, but can hardly fail to annoy the casual reader.

Sir Arthur Conan Doyle The Problem of Thor Bridge

Historical writing is a difficult genre since it is based on evidence which is always incomplete (and often inadequate), and only through conjecture and interpretation can the historian link the scattered pieces of evidence into a coherent narrative.\footnote{Following Ricoeur (1984: 138ff) all historical writing, whether describing 'events' or 'structures' across time, can be seen as taking the form of a narrative, i.e. representing a coherent description which we can follow 'as if' it were a story.} That this is true also for language history was emphasised by Lass (1997a: 4-5) whose notion of 'myth' largely agrees with the way the term 'narrative' is used here, that is, 'a myth in the widest sense is a story or image that structures some epistemic field'.

The histories of languages (as objects available to or made by linguists) are, like all histories, myths. We do have documents for portions of many of our histories; but even these are subject, like scripture, to exegesis: we don't know what they mean (the less, the older they are). We do however tell (and believe) stories about them, not just the documents but the languages they supposedly reflect [...].

Thus, rather than a 'solution' or definite answer to the question *wie ist es nun eigentlich gewesen?* (to rephrase Ranke's well-known statement) historical knowledge is best seen as 'a balance of probabilities' (Stanford 1994: 112); probabilities which can amount to near certainty, yet this certainty has no other basis than the interpretation of an always incomplete record. To emphasise the tentative nature of historical knowledge is not to open the doors to a postmodern 'anything goes'-relativism, since our interpretations can be measured against the criteria of empirical responsibility, coherence and plausibility (Lass 1997a: *ibid.*).

As an exercise in historical sociolinguistics this study was concerned not only with linguistic variation but also with the notion of speaker agency, a topic which has surfaced time and again in epistemological discussions of historical linguistics. Speaker agency was understood in agreement with Giddens from a third-person perspective, i.e. what matters is what speakers (as individuals, not groups) do (say or write) not what they intend. Motivations and intentions, on the other hand, enter our narratives not through the description of hypothetical intention-belief-action sequences *à la* Von Wright, but through the reconstruction of the speakers'
(historically and socially) situated knowledge about the world. Metalinguistic commentary grants us access to the speakers' discursively articulated knowledge, while their tacit knowledge (i.e. what they know about the use of language in social interaction, but cannot express verbally) can be reconstructed through close ethnographic analysis, carefully describing the speakers' use of language in social contexts. In contrast to the hermeneutical perspective, the individual's capacity for agency is seen as being bound by the unacknowledged conditions of action (rules and resources) and the unintended consequences of action. Macro-phenomena (social institutions or systems, the ex-planandum of social enquiry) are believed to emerge through regularities of conduct enacted by individual actors, that is, through social praxis. To describe such collective practices well-defined samples are necessary, and it was argued in § 1.3 that data requirements often constrain the application of a sociolinguistic perspective to historical data. The discussion of Giddens' work in § 1.2.3 was not meant to suggest that structuration theory (or any other social theory) should be imported into linguistic theorising as an explanatory deus ex machina, integrating our diverse sociolinguistic findings into one comprehensive theoretical framework. I agree with Horvath (1998: 448) that 'social theories and linguistic theories only partially overlap in the goals they set themselves and our understanding of both language and society is better served if the two remain true to their own goals and meet occasionally on some middle ground'. As historical linguists our main concern remains the description of change in the structural parts of language, and while social theory provides us with indispensable 'thinking tools', it does not solve any of our field-internal problems.

The patterns of variation observed in the corpus (1880-1922) indicate that for the acrolectal and possibly (upper) mesolectal varieties the process of language change at the Cape was slow and gradual, continuing in some cases (as shown in detail in the re-analysis of Conradie, § 6.1.2) patterns described for the late 18th century. The existence of such continuities suggests a process of slow change (stable variation is unlikely since the overall percentages were found to increase), extending into the late 19th and possibly early 20th centuries, which challenges the conventional dating of the emergence of Afrikaans to the late 18th century, and suggests that language change at the Cape proceeded at different speeds in the different social, ethnic and geographical varieties (as has been suggested by Roberge, see § 2.2.1). The acrolectal and mesolectal varieties described in this study developed in a complex process which balanced evolutive language change, endogenous innovation and contact-induced restructuring. Since the Cape colonial society was not strictly segregated until the late 19th century, contact between speakers of the different varieties of CDV (some of which were strongly marked by substratum influence and the universals of untutored second language acquisition) was an important mechanism of language change, leading to the transfer of substratum features as well as convergent hybridisation (see § 2.2.3). As argued by Roberge (1995: 81) the 'degree of influence
from one code on the other' was a function of the social standing as well as the geographical location (crudely: Cape Town and the Boland vs. the eastern and northern frontier) of the speakers. It is possible that basilectal varieties stabilised within relatively autonomous communities such as the Griquas and other non-dominant but cohesive groups (such as the Malay/Muslim community, see § 4.3.1).

A central argument of this study was that language change at the Cape did not result in the development of a rigid diglossic structure with Dutch (and its acrolectal varieties) as H, and the basilectal contact variety as L, each being used in discrete and functionally separate domains (see § 3.4 and § 8.1.4). While from a social point of view most varieties of CDV were located towards the 'low' end of the scale, stratification (social, geographic and ethnic) was complex, and followed patterns predicted by modern sociolinguistic theory. Rejecting the conventional interpretation of the Cape speech community as 'classically' diglossic also means that the written data should not be seen as reflecting a different type of language (H) which never occurred in spoken interaction, but rather as reflecting a different style ('formal'). In other words, while under diglossic conditions the relationship between spoken and written language is discontinuous, continuity along an informal-formal continuum is characteristic of social dialects. The interpretation of the Cape speech community as dialectal rather than diglossic is supported by the patterns of structured heterogeneity (phonological, grammatical and lexical) which were found to exist in the corpus (§8.1.1) - patterns typical for standard-with-dialects scenarios. I consider as unjustified and epistemologically problematic the argumentum ex silentio which is implicit in the work of Scholtz, Raidt and more recently Ponelis, i.e. the hypothesis that Afrikaans existed as a relatively uniform spoken vernacular from about 1800, yet diglossia prevented the structures of the spoken language (L) from surfacing in the written record (H). While the notion of diglossia fails to capture the complex nature of CDV variation, diglossic tendencies nevertheless existed, and certain groups within the speech community (typically members of the intelligentsia) described variation in terms of a clear H/L distinction, and perceived the two codes as highly focused and clear-cut entities (§ 3.3/3.4). In the corpus a small number of individuals were found to use 'Afrikaans' and 'Dutch' in strict alternation, and their texts could be identified unambiguously as representing one code or the other. Those who engaged in such patterns of code alternation (across and within texts) were young (born after 1865) and generally well educated, which agrees with Wright's (1991) observation that diglossia is not a naturally developing characteristic of speech communities, but emerges as a result of more or less conscious 'engineering' and is reproduced via the education system.

I furthermore argued that the 19th century dialect literature, which has conventionally been interpreted as being representative of the spoken language, is best understood as an example of 'variety imitation', characterised by the overgeneralisation of features which were still highly
variable in many varieties of the spoken language. The dialect literature as well as the linguistic nationalism of the GRA was important in that it helped to establish a typological conception of what constituted 'Afrikaans', and thus provided a relatively uniform linguistic model which may have influenced the linguistic practices of those, who by the late 1880s were found to use 'Afrikaans' (in alternation with Dutch) in their correspondence. At the same time linguistic practices not in line with the new model were identified by an emerging metalinguistic discourse (which can be described as an example of a 'standard language ideology') as unauthentic and undesirable. From the perspective of structuration theory one could argue that within a nationalist framework, language became an object of discursively motivated action: certain sociolinguistic 'rules' were described overtly as characteristic of the Afrikaner nation, and slowly became part of the discursive knowledge of (certain groups of) speakers, ultimately influencing aspects of their behaviour. If the sociohistorical analysis presented in this study is plausible, it follows that the adoption of Afrikaans as a new standard led to different processes of variety shifting, i.e. the bidialectal speakers shifted from one relatively well-defined linguistic system to another, others shifted (probably gradually) from a highly variable mesolectal variety of CDV to the highly regularised new standard language which had replaced the old Dutch standard in education and formal registers.

Absent from the corpus is the 'true' basilect (or rather the basilectal varieties). Focusing in the linguistic analysis on acro- and mesolectal varieties should not be interpreted as a 'failure to reach the vernacular' (Milroy 1979: 91 in his critique of Macaulay 1977) since basilects are not in any theoretical sense superior to other linguistic varieties; a point emphasised by Macaulay (1988: 110, 113) in his response to Milroy's critique:

What troubled me was that the criticism implied that the samples of speech recorded in the interviews were somehow not legitimate examples of Glasgow speech [...] I believe that they were all speaking the vernacular in Illich's sense of the language used by 'people who mean what they say and say what they mean in the context of everyday life' [...]. I agree that none of them was using the basilect in its most casual style; but for most of the respondents the basilect would not be the most casual form of speech. [...] The search for a pure vernacular is potentially dangerous, if it leads to the undervaluing of other varieties and a consequent lack of attention to them. The aim of sociolinguistic investigation should be to describe the totality of speech use in the community.

Similarly Mitchell (1982: 125) remarked in his discussion of Educated Spoken Arabic that:

Linguists have always found it fairly easy to describe vernaculars but have always resorted when doing so to an unconfessed purism, editing out without acknowledgement the prestigious 'literary' cum vernacular forms of the language that are in fact its commonest manifestations. [my emphasis]

By classifying texts which conformed neither to the emerging norms of Afrikaans nor to
the norms of continental Dutch as 'approximate Dutch' and the theoretically uninteresting ad hoc result of linguistic insecurity, Afrikaans historical linguists have engaged in a similar purism, and neglected those documents which reflect the structured but highly variable nature of many CDV varieties.

To further investigate the sociolinguistics of the Cape speech community, formal letters (ideally by the same individuals) should be added to the corpus. A formal letter was located for Hendrik van Huyssteen (addressed to C. Searle 7/7/1900; SAL-MSC 50-6), and while Hendrik van Huyssteen can be classified as mesolectal in his private letters (cluster II), the formal letter is written in acceptable Dutch, which supports Roberge's (1994a) hypothesis that depending on the function of the texts writers were able to screen out variation. Another avenue for research could be to conduct 'life history' interviews with individuals who grew up in the 1920s and 1930s when standard Afrikaans was implemented as an official language and medium of instruction, thus accelerating possible language variety shifts. Oral history as an ancillary technique of historical study has expanded rapidly since the 1960s, and although serious concerns have been raised about the accuracy of memory and the possible intrusion of subjective bias, it is generally accepted that oral sources (if collected through carefully structured interviews) can provide important historical evidence (for an overview of the limits and possibilities of oral history see Lummis 1998 [1983]). While sociohistorical interviews will tell us little about the more narrowly linguistic aspects, such interviews can be useful for the reconstruction of the sociolinguistic context, including language attitudes, perceptions of linguistic variation and domains of usage (discursive knowledge). Following the methodological practice of traditional historical linguistics and dialectology, access to older, basilectal features might exist in non-standard varieties, in particular in the 'speech of non-mobile, less educated and especially older speakers of these varieties' (see Roberge 1995: 76), and their analysis is thus important for a better understanding of the complex basilectal input. A more comprehensive reconstruction of the linguistic practices and changes at the Cape could thus be gained from a combination of sources as summarised in the diagram below (which is based loosely on Bailey & Ross 1992). The distinction between primary and secondary evidence roughly parallels the distinction 'unconsciously' - 'consciously' produced. That there is some value in categorising evidence in terms of 'consciousness' has been acknowledged by historians who have argued that what I have termed 'secondary evidence' is often deliberately biased, and therefore necessitates careful

\[2\] In (non-structured) exploratory interviews topics such as letter writing and reading, education and church elicited interesting comments. People also volunteered information about family members who had shifted from Dutch/Afrikaans to English, or who had been involved in the language movement after 1900.
interpretation (Stanford 1994: 143).³

Language practices and change at the Cape (around 1900)

primary evidence  secondary evidence
written oral written oral
private papers official papers evidence from contemporary dialect literature linguistic descriptions life history
(informal) (formal) non-standard varieties and metalinguistic commentary interviews

Apart from collecting and analysing new data which allowed a reconsideration of important aspects of the conventional historiography of Afrikaans, the topics discussed in this study have relevance for sociolinguistic theorising in general. Like other studies in historical sociolinguistics (Romaine 1982a, Devitt 1989, Lippi-Green 1994) the linguistic analysis summarised in chapters six, seven and eight has shown that written language exhibits the same kind of constraints and patterns of variation known from studies of spoken language. Variationist studies of written data can, therefore, help us to clarify much of the apparent disorder and randomness notable in the written record. As regards the process of linguistic standardisation, the patterns of variation discussed in chapter eight suggest that established standard varieties and newly focusing varieties have different structures of speaker membership. Established standard varieties show a Kernlandschaft or nucleus of individuals who are very similar in their linguistic behaviour; variation is largely the result of outliers or residuals, i.e. individual cases which show more marginal adherence to standard norms but which do not cluster themselves. Newly focused or focusing varieties combine variability with increasing uniformity, leading to a pattern in which individuals show relatively similar linguistic behaviour, without yet constituting a tight nucleus of very similar cases.

The social make-up of the 'new standard cluster' illustrates the path through which modern standard languages are typically diffused, i.e. via the socially (and geographically) mobile new professional class whose social networks are characterised by a great number of 'weak' ties (§ 4.3.2). These individuals resemble in a symbolic way the Milroys' concept of 'early adopters', i.e. individuals which are exposed to innovations by their relatively high number of weak ties,

³ Conventionally the distinction between primary and secondary sources is between evidence which is contemporary with the period in question (primary sources), and evidence which was produced after the period in question (secondary sources). In the diagram this is reflected in the distinction between 'written' (contemporary) and 'oral' (produced after the period in question) evidence.
but who are at the same time deeply integrated into their respective networks, and thus provide an influential model for other members of the network. Following Anderson's (1991) argument that nations are 'imagined communities' whose members (although not directly linked) have a strong feeling of belonging which is created and reinforced through nationalistic discourse, nationalist ideologies can be said to transform the weak ties which characterise modern societies (and which impede the establishment of new norms) into symbolically strong ties, thus facilitating the diffusion of new norms through the community.

The statistical methodology used in the analysis has emphasised descriptive methods. Pattern recognition rather than issues of statistical significance (although important and adding strength to an argument) seem to me to lie at the heart of variation analysis. The emphasis on multivariate analysis, and the assumption that dialects can be grouped together on the basis of some measure of similarity does not attempt to reintroduce the 'dialect myth' (i.e. the assumption that at some level we can find homogenous dialects; Harris 1990), rather the 'varieties' identified in chapter 8 must be seen as loose or fuzzy clusters, characterised by degrees of membership. The descriptive methodology does, furthermore, not imply that varieties or dialects (whether fuzzy or crisp) 'exist' independently from the linguist's description of them; they are first and foremost epistemological constructs.

In many ways this study is an extended argument against the principles of Aristotelian essentialism which have dominated much of western science. Under an essentialist ontology taxonomic categories are believed to have 'real' essences and to be discrete (i.e. each taxon is characterised by a unique combination of features). As a result objects can only belong to one taxon at a time, that is, the proposition 'X is a member of Z' must be either true or false; a rule summarised as tertium non datur, the principle of the excluded middle. Eager to establish 'language as a well-defined object' (Saussure 1966 [1922]: 14) linguistic theory has been marked by segregational thinking, trying to capture the essence or very nature of its object ('language') by 'postulating a thing where in reality there is a diversity of forms of behaviour and knowledge' (Joseph 1997: 10; the term 'segregational linguistics' goes back to Harris 1987). Essentialism has surfaced in linguistic thought also in the form of taxonomic dichotomies (competence-performance, synchrony-diachrony, natural-artificial, prescriptive-descriptive, standard-vernacular, etc.), or (more marginally) trilogies (acro-, meso-, basilect; Old, Middle, Modern Dutch; see Lass 1997b). Linguists (as well as other scientists) have often adopted an instrumentalist (nominalist) stance towards such taxonomies, and interpreted them as

---

4 To be fair, although Aristotle formulated the Law of Contradiction (X cannot be Z and not Z) and The Law of the Excluded Middle (X must be either Z or not Z) in his *Metaphysics* (1966: 58-59, 70), he never denied that truth can come in degrees: 'The more and the less are still present in the nature of things' (*ibid.*: 64).
epistemological conventions which structure the domains of enquiry (and as such taxonomies are akin to narratives or myths as mentioned above), but which are not necessarily 'real' (Dean 1979).

Whether seen as realities or methodological conventions, discrete taxonomies are at odds with the continuity which characterises complex systems. That all existence is intrinsically continuous was emphasised by Locke in his Essay on Human Understanding ('we shall find everywhere, that the several species are linked together, and differ but in almost insensible degrees', 1924 [1690]: 248), and in the 20th century C.S. Peirce noted that '[v]agueness is no more to be done away with in the world of logic than friction in mechanics' (cited in Nadin 1983). Lotfi Zadeh, the founder of fuzzy set theory, described the correlation of complexity and vagueness formally in his Law of Incompatibility:

As the complexity of a system increases, our ability to make precise yet significant statements about its behaviour diminishes until a threshold is reached beyond which precision and significance (or relevance) become almost mutually exclusive characteristics. (1973: 28)

Hence vagueness, that is, ill-defined boundaries and an overlapping of categories, is not a mark of faulty thinking nor a transitional state resulting from insufficient information (which one could address in terms of probability), but is said to define the world an sich.\(^5\) Psychological research has shown that not only do we have the ability to process vague (or fuzzy) categories, but that the very principles underlying human categorising are fuzzy, and cannot be reflected by a two-fold logic. Category membership is usually not assigned on the basis of necessary or sufficient attributes of objects, that is, it does not follow from the learning or knowledge of definitional rules (such as 'a square has four straight lines of equal length joined at right angles'), but rather depends on the perceived similarity of objects on a number of characteristic dimensions, and thus allows for degrees of category membership (see § 8.1.3).\(^6\) Rosch and Mervis (1975) described the principles underlying human categorising in terms of family resemblances: categories are organised around a number of properties which their elements have in common, however, none of these properties defines category membership, and few (or no)

---

\(^5\) From a nominalist perspective continuity gives way to discreteness only at the level of the individual, i.e. while superordinate categories are generally ill-defined constructs, their individual members are discrete objects characterised by a unique combination of features.

\(^6\) The notion of similarity, although central to theories of cognition, is not unproblematic since in theory 'any two things share an arbitrary number of predicates and differ from each other in an arbitrary number of ways' (Medin, Goldstone & Genter 1993: 255). To constrain the predicates that enter into the computation of similarity is the aim of a position known as psychological essentialism. While psychological essentialism does not try to reintroduce the notion that things have 'real' essences, it suggests that categorisation (i.e. the perception of similarities) is constrained by the naïve and often essentialist theories people have about the world (Wattenmaker, Nakamura & Medin 1988, Medin & Ortony 1989).
elements in a category will exhibit all properties. The more properties an element shares with other elements, the more prototypical it is of the category; such prototypical members were shown to be learned earlier and be classified faster than more atypical members. The term *family resemblance* originates from Wittgenstein’s discussion of games in the *Philosophischen Untersuchungen* (1953: 66, 67):

Was ist allen diesen gemeinsam? - Sag nicht: "Es muß ihnen etwas gemeinsam sein, sonst hießen sie nicht 'Spiele'" - sondern *schau*, ob ihnen allen etwas gemeinsam ist. - Denn, wenn du sie anschaut, wirst du zwar nicht etwas sehen, was *allen* gemeinsam wäre, aber du wirst Ähnlichkeiten, Verwandtschaften sehen, und zwar ein ganze Reihe. Wie gesagt: denk nicht, sondern schau. [...] Und das Ergebnis dieser Betrachtungen lautet nun: Wir sehen ein kompliziertes Netz von Ähnlichkeiten, die einander übergreifen und kreuzen. [...] Ich kann diese Ähnlichkeiten nicht besser charakterisieren, als durch das Wort 'Familienähnlichkeiten'; denn so übergreifen und kreuzen sich die verschiedenen Ähnlichkeiten, die zwischen den Gliedern einer Familie bestehen: Wuchs, Gesichtszüge, Augenfarbe, Gang, Temperament, etc.etc. - Und ich werde sagen: die "Spiele" bilden eine Familie.

Relative similarities and differences are no stranger to scientific thought; they enter into our definition of languages vs. dialects (as reflected, for example, in Kloss’ term *Abstandsersprache*), as well as into the biologist’s classification of species (‘the amount of difference is one very important criterion when settling whether two forms should be ranked as species or varieties’, Darwin 1902 [1859]: 52).

Considering (i) the observation that complex systems are non-discrete and therefore cannot be described meaningfully by using precise categories (Zadeh’s Law of Incompatibility), and (ii) the results of cognition research which indicate that we are perfectly able to comprehend and use fuzzy categories in everyday life, the question is whether we help or hinder our academic projects by describing our fields of enquiry in terms of discrete and well-defined categories and taxonomies.7 Wittgenstein, turning against Frege’s two-valued logic, argued for the need of ill-defined concepts, and the importance of prototypes and examples in descriptive and explanatory statements (1953: 71, 75, 77).


---

7 Whether crisp or fuzzy our scientific categories are, of course, explanatory models, and do not necessarily reflect the properties of ‘reality’. I agree with Romaine (1982a: 283-284) in that ‘it is a mistake to put questions about the reality or existence of theoretical entities into the center of linguistic theory. Belief in a theory need not entail essentialism or realism, but only a commitment to a belief that a phenomenon can be described by the application of a theory.’
The need for fuzzy-edged scientific constructs and taxonomies surfaces whenever a seemingly crisp concept such as, for example, diglossia is applied to real-life situations, and linguists soon acknowledged that 'diglossia' is not a [+ ] feature, but that examples were characterised by a gradient of representativeness (or similarity) with respect to the four prototypes identified in Ferguson's original paper (see § 3.4). The same holds for other sociolinguistic or sociological concepts such as 'the working class', 'the bilingual individual', 'the speech community', 'pidgin/creoles', etc.; they remain marked by a certain amount of vagueness and flexibility which makes them resistant to crisp essentialist definitions. While it is important to recognise the elasticity of the boundaries which define our descriptive categories and to accept the ill-defined ontology of complex systems, to deny the value of crisp categories would be to throw out the baby with the bathwater. Well-defined categories allow us to create 'islands of meaning' since it is their very boundaries 'that allow us to perceive any "thing" at all' (Zerubavel 1991: 118). They certainly don't 'carve nature at its joints' as suggested by Plato, since there are multiple ways of carving up the world; yet the lines we draw can help us to develop perspectives which make it possible to perceive in a reductionist fashion certain aspects of the ill-defined object 'language'. They allow us to tell macro-stories (Lass 1997a: 288) about language as well as about the world in general by purposefully ignoring its complexities.
APPENDICES
APPENDIX A: STATISTICA - A Short Description of the Computer Package

The computer package used in this study is STATISTICA 1984-1995 (version 5), developed by StatSoft Inc. STATISTICA offers modules for several types of multivariate analysis including cluster analysis, principal components analysis and multidimensional scaling.

The cluster analysis module includes implementation of three clustering methods: k-means clustering, hierarchical clustering and two-way joining. Cases as well as variables can be clustered. An important feature of STATISTICA is that the program offers a variety of distance measures: Euclidian, Squared Euclidrian, Manhattan, Chebychev, power distances, percent disagreement, and correlation coefficients (1-Pearson r). A selection of linkage rules is also available to the researcher: single linkage, complete linkage, weighted and unweighted group averages, weighted and unweighted centroid, and Ward's method. The output of the cluster analysis module includes extended diagnostic statistics (complete amalgamation schedule with cohesion levels for hierarchical clustering, the ANOVA-table for k-means clustering) and graphic options (tree-diagram, two-way joining plot, plot of means in k-means clustering and graph of amalgamation schedule).

The multidimensional scaling module performs non-metric scaling. Dissimilarity matrices with up to 90 objects can be analysed and a maximum of nine dimensions can be specified by the user. The starting configuration used in chapters 5 and 8 is Standard Guttman-Lingoes. The program performs monotone transformation of the input data to minimise the stress value and the coefficient of alienation. The output options include: two- and three dimensional graphs of the reproduced configuration, Shepard's diagram and a numeric summary of the final configuration.

The principal components analysis module in STATISTICA can analyse matrices with up to 300 variables. Four different types of factor rotations are available: varimax, biorthogonal, quartimax and equamax (for a detailed description of the rotation techniques see STATISTICA 1995 III: 3227-3228). The output of the program includes: regular and squared eigenvalues, scree plot, component loadings, communalities and component scores.
<table>
<thead>
<tr>
<th>name</th>
<th>year of birth</th>
<th>sex</th>
<th>class</th>
<th>ethnicity</th>
<th>place(s) of residence</th>
</tr>
</thead>
<tbody>
<tr>
<td>April, Aletta</td>
<td>1890</td>
<td>F</td>
<td>II</td>
<td>Coloured</td>
<td>Wellington</td>
</tr>
<tr>
<td>April, Anton</td>
<td>1869</td>
<td>M</td>
<td>IV</td>
<td>Coloured</td>
<td>Genadendal</td>
</tr>
<tr>
<td>Baalie, Fredrik</td>
<td>1864</td>
<td>M</td>
<td>II</td>
<td>Coloured</td>
<td>Genadendal</td>
</tr>
<tr>
<td>Baalie, Rudolf</td>
<td>1854</td>
<td>M</td>
<td>II</td>
<td>Coloured</td>
<td>Wellington</td>
</tr>
<tr>
<td>Basson, Mary</td>
<td>1860</td>
<td>F</td>
<td>III</td>
<td>White</td>
<td>Wellington</td>
</tr>
<tr>
<td>Baartman, Susie</td>
<td>1875</td>
<td>F</td>
<td>III</td>
<td>White</td>
<td>Wellington</td>
</tr>
<tr>
<td>Beyers, Catharina F.</td>
<td>1885</td>
<td>F</td>
<td>III</td>
<td>White</td>
<td>Wellington</td>
</tr>
<tr>
<td>Blake, Ebenzer J.</td>
<td>1854</td>
<td>M</td>
<td>IV</td>
<td>White</td>
<td>Wellington</td>
</tr>
<tr>
<td>Bonthuis, Jan</td>
<td>1830</td>
<td>M</td>
<td>III</td>
<td>White</td>
<td>Paarl</td>
</tr>
<tr>
<td>Bosman, Louisa</td>
<td>1885</td>
<td>F</td>
<td>IV</td>
<td>Coloured</td>
<td>Wellington</td>
</tr>
<tr>
<td>Botha, Louis</td>
<td>1862</td>
<td>M</td>
<td>III</td>
<td>White</td>
<td>OFS</td>
</tr>
<tr>
<td>Breuris, Frans M.C.</td>
<td>1856</td>
<td>M</td>
<td>III</td>
<td>White</td>
<td>Albertinia</td>
</tr>
<tr>
<td>Bruin, Barend J.</td>
<td>1841</td>
<td>M</td>
<td>III</td>
<td>White</td>
<td>Lady Grey</td>
</tr>
<tr>
<td>Bruin, Johanna</td>
<td>1877</td>
<td>F</td>
<td>II</td>
<td>White</td>
<td>Lady Grey, Cape Town</td>
</tr>
<tr>
<td>Bruin, Johann J.</td>
<td>1864</td>
<td>M</td>
<td>III</td>
<td>White</td>
<td>Barkley East</td>
</tr>
<tr>
<td>Brümmel, Nikolaas J.</td>
<td>1876</td>
<td>M</td>
<td>II</td>
<td>White</td>
<td>Lady Grey, Cape Town, Stellenbosch</td>
</tr>
<tr>
<td>Burger, Jacobus A.</td>
<td>1831</td>
<td>M</td>
<td>III</td>
<td>White</td>
<td>Montagu</td>
</tr>
<tr>
<td>Cillie, Francois S.</td>
<td>1848</td>
<td>M</td>
<td>III</td>
<td>White</td>
<td>Paarl</td>
</tr>
<tr>
<td>Coopman, Cornelius</td>
<td>1856</td>
<td>M</td>
<td>IV</td>
<td>Coloured</td>
<td>Hanover</td>
</tr>
<tr>
<td>Cronje, Daniel C.</td>
<td>1854</td>
<td>M</td>
<td>III</td>
<td>White</td>
<td>Albertinia</td>
</tr>
<tr>
<td>Cronje, Hendrik S.</td>
<td>1848</td>
<td>M</td>
<td>III</td>
<td>White</td>
<td>Albertinia</td>
</tr>
<tr>
<td>Cronje, Maria</td>
<td>1885</td>
<td>F</td>
<td>II</td>
<td>White</td>
<td>Albertinia</td>
</tr>
<tr>
<td>Cronje, Wessels C.</td>
<td>1860</td>
<td>M</td>
<td>III</td>
<td>White</td>
<td>Albertinia</td>
</tr>
<tr>
<td>Dampies, Piet</td>
<td>1860</td>
<td>M</td>
<td>IV</td>
<td>Coloured</td>
<td>Albertinia</td>
</tr>
<tr>
<td>Daniels, Sophie</td>
<td>1862</td>
<td>F</td>
<td>IV</td>
<td>Coloured</td>
<td>Cape Town</td>
</tr>
<tr>
<td>De Villiers, Jacobus De W.</td>
<td>1866</td>
<td>M</td>
<td>III</td>
<td>White</td>
<td>Paarl</td>
</tr>
<tr>
<td>Name</td>
<td>Year of Birth</td>
<td>Sex</td>
<td>Class</td>
<td>Ethnicity</td>
<td>Place(s) of Residence</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------------</td>
<td>-----</td>
<td>-------</td>
<td>-----------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>DeVilliers, J.H.</td>
<td>1842</td>
<td>M</td>
<td>I</td>
<td>White</td>
<td>Cape Town</td>
</tr>
<tr>
<td>DeVilliers, Jan J.</td>
<td>1856</td>
<td>M</td>
<td>III</td>
<td>White</td>
<td>Kimberly</td>
</tr>
<tr>
<td>DeVos, Helena Cath.</td>
<td>1848</td>
<td>F</td>
<td>III</td>
<td>White</td>
<td>Kimberly</td>
</tr>
<tr>
<td>DeVries, Anna</td>
<td>1850</td>
<td>F</td>
<td>II</td>
<td>White</td>
<td>Swayoodam</td>
</tr>
<tr>
<td>DeWaal, Beatrice</td>
<td>1860</td>
<td>F</td>
<td>III</td>
<td>White</td>
<td>Wellington</td>
</tr>
<tr>
<td>DeWaal, Jan H.H.</td>
<td>1871</td>
<td>M</td>
<td>II</td>
<td>White</td>
<td>Stellenbosch, Cape Town</td>
</tr>
<tr>
<td>DeWet, D.C.</td>
<td>1875</td>
<td>M</td>
<td>II</td>
<td>White</td>
<td>Cape Town</td>
</tr>
<tr>
<td>DeWet, Nicolaas J.</td>
<td>1870</td>
<td>M</td>
<td>II</td>
<td>White</td>
<td>Cape Town</td>
</tr>
<tr>
<td>Dippenaar, Johannes</td>
<td>1875</td>
<td>M</td>
<td>III</td>
<td>White</td>
<td>NA</td>
</tr>
<tr>
<td>DuPreez, Johannes G.</td>
<td>1863</td>
<td>M</td>
<td>III</td>
<td>White</td>
<td>Oudtshoorn</td>
</tr>
<tr>
<td>DuPreez, P.E.</td>
<td>1860</td>
<td>M</td>
<td>III</td>
<td>White</td>
<td>Burghersdorp</td>
</tr>
<tr>
<td>Eksteen, Elizabeth J.H.</td>
<td>1860</td>
<td>F</td>
<td>III</td>
<td>White</td>
<td>Wellington, Boshoff (OFS)</td>
</tr>
<tr>
<td>Eksteen, P.J.</td>
<td>1875</td>
<td>M</td>
<td>IV</td>
<td>Coloured</td>
<td>Genadendal, Cape Town</td>
</tr>
<tr>
<td>Es(s)op, Napie</td>
<td>1885</td>
<td>F</td>
<td>IV</td>
<td>Coloured</td>
<td>Worcester</td>
</tr>
<tr>
<td>Furter, Engela J.M.</td>
<td>1865</td>
<td>F</td>
<td>III</td>
<td>White</td>
<td>Wellington, Robben Island</td>
</tr>
<tr>
<td>Goosen, Gideon J.C.</td>
<td>1884</td>
<td>M</td>
<td>III</td>
<td>White</td>
<td>Wellington</td>
</tr>
<tr>
<td>Goosen, Jan A.B.</td>
<td>1887</td>
<td>M</td>
<td>III</td>
<td>White</td>
<td>Wellington</td>
</tr>
<tr>
<td>Goosen, Pieter A. jun.</td>
<td>1885</td>
<td>M</td>
<td>II</td>
<td>White</td>
<td>Wellington</td>
</tr>
<tr>
<td>Goosen, Pieter A. sen.</td>
<td>1862</td>
<td>M</td>
<td>III</td>
<td>White</td>
<td>Wellington, Bethal</td>
</tr>
<tr>
<td>Goosen, Sebastina</td>
<td>1890</td>
<td>F</td>
<td>II</td>
<td>White</td>
<td>Wellington</td>
</tr>
<tr>
<td>Goosen, Susanna J.F. sen.</td>
<td>1865</td>
<td>F</td>
<td>III</td>
<td>White</td>
<td>Wellington</td>
</tr>
<tr>
<td>Goosen, Susanna J.F. jun.</td>
<td>1885</td>
<td>F</td>
<td>II</td>
<td>White</td>
<td>Bloemfontein, Stellenbosch</td>
</tr>
<tr>
<td>Grosskopf, J.F.W.</td>
<td>1885</td>
<td>M</td>
<td>II</td>
<td>White</td>
<td>Wellington, Transvaal</td>
</tr>
<tr>
<td>Hauptfleisch, Andries P.</td>
<td>1880</td>
<td>M</td>
<td>III</td>
<td>White</td>
<td>Wellington, Transvaal</td>
</tr>
<tr>
<td>Hatting, Benjamin</td>
<td>1829</td>
<td>M</td>
<td>IV</td>
<td>White</td>
<td>Prins Albert</td>
</tr>
<tr>
<td>Henderson, James</td>
<td>1875</td>
<td>M</td>
<td>IV</td>
<td>White</td>
<td>Cape Town</td>
</tr>
<tr>
<td>Hendrikz, Sophie</td>
<td>1875</td>
<td>F</td>
<td>III</td>
<td>White</td>
<td>Somerset West, Cape Town</td>
</tr>
<tr>
<td>Heroldt, Michiel J.</td>
<td>1819</td>
<td>M</td>
<td>III</td>
<td>White</td>
<td>Graaff Reinet</td>
</tr>
<tr>
<td>Hillebrand, Martha C.</td>
<td>1820</td>
<td>F</td>
<td>III</td>
<td>White</td>
<td>Riversdale, Knysna</td>
</tr>
</tbody>
</table>

316
<table>
<thead>
<tr>
<th>name</th>
<th>year of birth</th>
<th>sex</th>
<th>class</th>
<th>ethnicity</th>
<th>place(s) of residence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hlela, Johannes</td>
<td>1860</td>
<td>M</td>
<td>II</td>
<td>Black</td>
<td>Genadendal, Queenstown</td>
</tr>
<tr>
<td>Hoogenhout, Dan. M.</td>
<td>1885</td>
<td>M</td>
<td>II</td>
<td>White</td>
<td>Wellington, Germiston</td>
</tr>
<tr>
<td>Hoogenhout, Francois P.</td>
<td>1880</td>
<td>M</td>
<td>II</td>
<td>White</td>
<td>Wellington, Belfast</td>
</tr>
<tr>
<td>Hoogenhout, Hendrina</td>
<td>1885</td>
<td>F</td>
<td>II</td>
<td>White</td>
<td>Germiston</td>
</tr>
<tr>
<td>Hoogenhout, Hester</td>
<td>1885</td>
<td>F</td>
<td>II</td>
<td>White</td>
<td>Pretoria</td>
</tr>
<tr>
<td>Hoogenhout, Johanna</td>
<td>1877</td>
<td>F</td>
<td>II</td>
<td>White</td>
<td>Wellington</td>
</tr>
<tr>
<td>Hoogenhout, Nicolaas M.</td>
<td>1876</td>
<td>M</td>
<td>II</td>
<td>White</td>
<td>Wellington, Pretoria</td>
</tr>
<tr>
<td>Hoogenhout, Petrus I.</td>
<td>1884</td>
<td>M</td>
<td>II</td>
<td>White</td>
<td>Wellington, Bethal</td>
</tr>
<tr>
<td>Hoogenhout, Wynanda</td>
<td>1888</td>
<td>F</td>
<td>II</td>
<td>White</td>
<td>Wellington, Transvaal</td>
</tr>
<tr>
<td>January, Wilhelmina</td>
<td>1850</td>
<td>F</td>
<td>IV</td>
<td>Coloured</td>
<td>Elim</td>
</tr>
<tr>
<td>Koch, J.A.</td>
<td>1875</td>
<td>M</td>
<td>II</td>
<td>White</td>
<td>Albertinia</td>
</tr>
<tr>
<td>Kotze, T.N.</td>
<td>1845</td>
<td>M</td>
<td>III</td>
<td>White</td>
<td>Upington</td>
</tr>
<tr>
<td>Krige, Leopold J.</td>
<td>1884</td>
<td>M</td>
<td>II</td>
<td>White</td>
<td>Stellenbosch</td>
</tr>
<tr>
<td>Kuhn, Pieter G.</td>
<td>1854</td>
<td>M</td>
<td>III</td>
<td>White</td>
<td>Swellendam, Worcester</td>
</tr>
<tr>
<td>Laubscher, Mimi</td>
<td>1885</td>
<td>F</td>
<td>II</td>
<td>White</td>
<td>Stellenbosch</td>
</tr>
<tr>
<td>Leipoldt, Carolina J.M.</td>
<td>1808</td>
<td>F</td>
<td>II</td>
<td>White</td>
<td>Cape Town, Wupperthal</td>
</tr>
<tr>
<td>leRoux, D.J.</td>
<td>1820</td>
<td>F</td>
<td>III</td>
<td>White</td>
<td>Wellington</td>
</tr>
<tr>
<td>leRoux, Petrus J.</td>
<td>1875</td>
<td>M</td>
<td>III</td>
<td>White</td>
<td>Wellington</td>
</tr>
<tr>
<td>Louw, Andries A. sen.</td>
<td>1827</td>
<td>M</td>
<td>II</td>
<td>White</td>
<td>OFS, Murraysburg, Paarl</td>
</tr>
<tr>
<td>Louw, Andries A. jun.</td>
<td>1862</td>
<td>M</td>
<td>II</td>
<td>White</td>
<td>Paarl, 'Mashonaland'</td>
</tr>
<tr>
<td>Louw, Francina S.</td>
<td>1872</td>
<td>F</td>
<td>II</td>
<td>White</td>
<td>Riebeek West, 'Mashonaland'</td>
</tr>
<tr>
<td>Louw, Maria J.</td>
<td>1856</td>
<td>F</td>
<td>II</td>
<td>White</td>
<td>OFS, Murraysburg, Cape Town</td>
</tr>
<tr>
<td>Louw, Rijkie Hester</td>
<td>1875</td>
<td>F</td>
<td>II</td>
<td>White</td>
<td>Cape Town</td>
</tr>
<tr>
<td>Louwrens, Anna S.</td>
<td>1817</td>
<td>F</td>
<td>III</td>
<td>White</td>
<td>Albertinia</td>
</tr>
<tr>
<td>Lourens, Joh.F.</td>
<td>1877</td>
<td>M</td>
<td>III</td>
<td>White</td>
<td>Albertinia</td>
</tr>
<tr>
<td>Malan, Daniel G.</td>
<td>1845</td>
<td>M</td>
<td>II</td>
<td>White</td>
<td>Paarl</td>
</tr>
<tr>
<td>Malan, Elizabeth J.</td>
<td>1848</td>
<td>F</td>
<td>III</td>
<td>White</td>
<td>Paarl</td>
</tr>
<tr>
<td>Malan, Francois S.</td>
<td>1871</td>
<td>M</td>
<td>II</td>
<td>White</td>
<td>Paarl, Cape Town</td>
</tr>
<tr>
<td>Malan, Jacques</td>
<td>1900</td>
<td>M</td>
<td>II</td>
<td>White</td>
<td>Cape Town</td>
</tr>
</tbody>
</table>

317
<table>
<thead>
<tr>
<th>name</th>
<th>year of birth</th>
<th>sex</th>
<th>class</th>
<th>ethnicity</th>
<th>place(s) of residence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malan, Leloa</td>
<td>1895</td>
<td>M</td>
<td>II</td>
<td>White</td>
<td>Wellington, Knysna</td>
</tr>
<tr>
<td>Malan, Maria, C.</td>
<td>1889</td>
<td>F</td>
<td>III</td>
<td>White</td>
<td>Paarl</td>
</tr>
<tr>
<td>Malherbe, C.</td>
<td>1895</td>
<td>F</td>
<td>II</td>
<td>White</td>
<td>Paarl</td>
</tr>
<tr>
<td>Marais, F.P.</td>
<td>1850</td>
<td>M</td>
<td>III</td>
<td>White</td>
<td>Stellenbosch</td>
</tr>
<tr>
<td>Marais, Jacobus P.Z.</td>
<td>1850</td>
<td>M</td>
<td>III</td>
<td>White</td>
<td>De Aar, Paarl</td>
</tr>
<tr>
<td>Marais, Helena Elizabeth</td>
<td>1850</td>
<td>F</td>
<td>III</td>
<td>White</td>
<td>Wellington</td>
</tr>
<tr>
<td>Marais, 'Rachie'</td>
<td>1885</td>
<td>F</td>
<td>II</td>
<td>White</td>
<td>Paarl</td>
</tr>
<tr>
<td>Marais, Stephanus Jacob.</td>
<td>1860</td>
<td>M</td>
<td>III</td>
<td>White</td>
<td>De Aar, Paarl</td>
</tr>
<tr>
<td>Marais, 'Maggie'</td>
<td>1895</td>
<td>M</td>
<td>III</td>
<td>White</td>
<td>Wellington, Bethal</td>
</tr>
<tr>
<td>Matodiana, Mozes</td>
<td>1875</td>
<td>M</td>
<td>II</td>
<td>Black</td>
<td>Genadendal, Port Elizabeth</td>
</tr>
<tr>
<td>Mlimkulu, Mozes</td>
<td>1875</td>
<td>M</td>
<td>II</td>
<td>Black</td>
<td>Genadendal, Port Elisabeth</td>
</tr>
<tr>
<td>Muller, Hermanus Steyn</td>
<td>1845</td>
<td>M</td>
<td>III</td>
<td>White</td>
<td>Albertinia</td>
</tr>
<tr>
<td>November, Jacobus</td>
<td>1895</td>
<td>M</td>
<td>IV</td>
<td>White</td>
<td>Wellington</td>
</tr>
<tr>
<td>Odendal, Johanna</td>
<td>1885</td>
<td>F</td>
<td>III</td>
<td>White</td>
<td>Swellendam</td>
</tr>
<tr>
<td>Oosthuizen, Michael A.M.</td>
<td>1864</td>
<td>F</td>
<td>III</td>
<td>White</td>
<td>Riversdale</td>
</tr>
<tr>
<td>Pauw, Anna</td>
<td>1900</td>
<td>F</td>
<td>II</td>
<td>White</td>
<td>Wellington</td>
</tr>
<tr>
<td>Pienaar, Hendrik H.</td>
<td>1840</td>
<td>M</td>
<td>III</td>
<td>White</td>
<td>Albertinia</td>
</tr>
<tr>
<td>Roos, 'Lenie'</td>
<td>1885</td>
<td>F</td>
<td>IV</td>
<td>White</td>
<td>OFS, Somerset West</td>
</tr>
<tr>
<td>Roussouw, Aletta M.</td>
<td>1850</td>
<td>F</td>
<td>III</td>
<td>White</td>
<td>Wellington</td>
</tr>
<tr>
<td>Roux, Katie</td>
<td>1895</td>
<td>F</td>
<td>IV</td>
<td>White</td>
<td>Hawston</td>
</tr>
<tr>
<td>Schabort, Elizabeth A.S.</td>
<td>1871</td>
<td>F</td>
<td>III</td>
<td>White</td>
<td>Bethlehem (OFS)</td>
</tr>
<tr>
<td>Schrade, Hester</td>
<td>1890</td>
<td>F</td>
<td>II</td>
<td>White</td>
<td>Griekwastad</td>
</tr>
<tr>
<td>Small, Nicolaas</td>
<td>1885</td>
<td>M</td>
<td>IV</td>
<td>Coloured</td>
<td>Gordon's Bay</td>
</tr>
<tr>
<td>Smit, 'Bettie'</td>
<td>1890</td>
<td>F</td>
<td>II</td>
<td>White</td>
<td>Clanwilliam</td>
</tr>
<tr>
<td>Smith, Edmund T.</td>
<td>1873</td>
<td>M</td>
<td>III</td>
<td>White</td>
<td>Barkley East</td>
</tr>
<tr>
<td>Smuts, J.A. ('Koos')</td>
<td>1861</td>
<td>M</td>
<td>III</td>
<td>White</td>
<td>Malmesbury</td>
</tr>
<tr>
<td>Smuts, J.C.</td>
<td>1870</td>
<td>M</td>
<td>I</td>
<td>White</td>
<td>Albertinia</td>
</tr>
<tr>
<td>Smuts, Sybella M. ('Isie')</td>
<td>1870</td>
<td>F</td>
<td>I</td>
<td>White</td>
<td>Albertinia</td>
</tr>
<tr>
<td>Stanford, Cornelia</td>
<td>1869</td>
<td>F</td>
<td>III</td>
<td>White</td>
<td>Swellendam</td>
</tr>
<tr>
<td>Steyn, Anna</td>
<td>1875</td>
<td>F</td>
<td>III</td>
<td>White</td>
<td>Swellendam</td>
</tr>
<tr>
<td>name</td>
<td>year of birth</td>
<td>sex</td>
<td>class</td>
<td>ethnicity</td>
<td>place(s) of residence</td>
</tr>
<tr>
<td>-----------------------</td>
<td>---------------</td>
<td>-----</td>
<td>-------</td>
<td>-----------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Steyn, Daniel</td>
<td>1830</td>
<td>M</td>
<td>III</td>
<td>White</td>
<td>Albertinia</td>
</tr>
<tr>
<td>Steyn, Elizabeth G.</td>
<td>?1860</td>
<td>F</td>
<td>III</td>
<td>White</td>
<td>Albertinia</td>
</tr>
<tr>
<td>Steyn, Rachel I.</td>
<td>1865</td>
<td>F</td>
<td>II</td>
<td>White</td>
<td></td>
</tr>
<tr>
<td>Uys, Susan</td>
<td>?1875</td>
<td>F</td>
<td>III</td>
<td>White</td>
<td>Albertinia</td>
</tr>
<tr>
<td>vd Merwe, C.J.</td>
<td>?1875</td>
<td>F</td>
<td>III</td>
<td>White</td>
<td>Ceres</td>
</tr>
<tr>
<td>vHuyssteen, A.J.Z. ('Janie')</td>
<td>1876</td>
<td>F</td>
<td>III</td>
<td>White</td>
<td>Knysna, Oudtshoorn</td>
</tr>
<tr>
<td>vHuyssteen, Anna Chr.</td>
<td>1873</td>
<td>F</td>
<td>III</td>
<td>White</td>
<td>Knysna</td>
</tr>
<tr>
<td>vHuyssteen, Dan. P.S.</td>
<td>1869</td>
<td>M</td>
<td>III</td>
<td>White</td>
<td>Knysna</td>
</tr>
<tr>
<td>vHuyssteen, Eliza J.</td>
<td>1882</td>
<td>F</td>
<td>II</td>
<td>White</td>
<td>Knysna</td>
</tr>
<tr>
<td>vHuyssteen, E.C.P. ('Petronella')</td>
<td>1854</td>
<td>F</td>
<td>III</td>
<td>White</td>
<td>Riversdale, Knysna</td>
</tr>
<tr>
<td>vHuyssteen, Hendrik</td>
<td>1844</td>
<td>M</td>
<td>III</td>
<td>White</td>
<td>Knysna</td>
</tr>
<tr>
<td>vHuyssteen, J.P.J.W. ('Algie')</td>
<td>1871</td>
<td>M</td>
<td>II</td>
<td>White</td>
<td>Knysna, Fauresmith (OFS)</td>
</tr>
<tr>
<td>vHuyssteen, J.H. ('Minnie')</td>
<td>?1885</td>
<td>F</td>
<td>III</td>
<td>White</td>
<td>Knysna</td>
</tr>
<tr>
<td>vHuyssteen, M.C. ('Katie')</td>
<td>1874</td>
<td>F</td>
<td>III</td>
<td>White</td>
<td>Knysna, Albertinia, Swellendam</td>
</tr>
<tr>
<td>vHuyssteen, Petrus H.H.</td>
<td>1872</td>
<td>M</td>
<td>II</td>
<td>White</td>
<td>Knysna</td>
</tr>
<tr>
<td>vHuyssteen, Matthys P.Z.</td>
<td>1881</td>
<td>M</td>
<td>III</td>
<td>White</td>
<td>Knysna</td>
</tr>
<tr>
<td>vOosthuizen, Lettie</td>
<td>?1885</td>
<td>F</td>
<td>III</td>
<td>White</td>
<td>Riversdale</td>
</tr>
<tr>
<td>vZyl, Susan</td>
<td>?1885</td>
<td>F</td>
<td>III</td>
<td>White</td>
<td>Albertinia</td>
</tr>
<tr>
<td>Vermaak, Hester S.</td>
<td>1864</td>
<td>F</td>
<td>III</td>
<td>White</td>
<td>Albertinia</td>
</tr>
<tr>
<td>'Hester'</td>
<td>?1875</td>
<td>F</td>
<td>III</td>
<td>White</td>
<td>Middelburg</td>
</tr>
<tr>
<td>'Gaybba'</td>
<td>?1875</td>
<td>F</td>
<td>IV</td>
<td>?Coloured</td>
<td>Wellington</td>
</tr>
<tr>
<td>'Auntie'</td>
<td>?1860</td>
<td>F</td>
<td>II</td>
<td>White</td>
<td>Pniel</td>
</tr>
</tbody>
</table>

Birthdates preceded by ? are approximates based on the general information collected, but no death notices were available for these individuals. Although there is room for error, with respect to the question whether individuals belong to age group 1 or age group 2, the estimates can be seen as sufficient.

Class membership was assigned on the basis of occupational and financial information obtained from death notices, the voters' role 1903, the South African directories as well as social information provided in the letters themselves.
<table>
<thead>
<tr>
<th>Name</th>
<th>Description of the Documents</th>
<th>No. of Words</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>April, Aletta</td>
<td>letters to C.P. Hoogenhout (27/1/1913, 19/9/1913, 25/10/1918, two letters c.1918)</td>
<td>672</td>
<td>SAL-MSC 1-26-4/ 1-41-8</td>
</tr>
<tr>
<td>April, Anton</td>
<td>letter to Mr. Schmidt (6/7/1913)</td>
<td>221</td>
<td>G- Vorsteheramt Briefe 1913-1914</td>
</tr>
<tr>
<td>Baalie, Fredrik</td>
<td>letter to P.O. Henning (3/12/1896)</td>
<td>157</td>
<td>G-brieue onderwyser aan Superintendent 1850-1900</td>
</tr>
<tr>
<td>Baalie, Rudolf</td>
<td>letters to P.O. Henning (23/2/1894, 30/10/1897)</td>
<td>639</td>
<td>G-brieue onderwyser aan Superintendent 1850-1900</td>
</tr>
<tr>
<td>Baartman Susie</td>
<td>letters to C.P. Hoogenhout (18/7/1895, 30/2/1902, 22/8/19, 1/3/1920)</td>
<td>605</td>
<td>SAL-MSC 1-25-5</td>
</tr>
<tr>
<td>Beyers, Catharina E.</td>
<td>letters to C.P. Hoogenhout (20/1/1906, 25/2/1906, one letter circa 1918)</td>
<td>1475</td>
<td>SAL-MSC 1-25-7</td>
</tr>
<tr>
<td>Bonhuis, Jan</td>
<td>letters to C.P. Hoogenhout (26/4/1909, 3/5/1909)</td>
<td>100</td>
<td>SAL-MSC 1-41-1</td>
</tr>
<tr>
<td>Bosman, Louis</td>
<td>letter to C.P. Hoogenhout (26/1/1900)</td>
<td>161</td>
<td>SAL-MSC 1-41-8</td>
</tr>
<tr>
<td>Breuris, Frans M.C.</td>
<td>letters to Katie v. Huyssteen (22/1/1906, 28/12/1905), to David Malan (11/12/1905)</td>
<td>910</td>
<td>SAL-MSC 853-4-1/ 853-1-3/ 852-1-13</td>
</tr>
<tr>
<td>Brümmer Johanna</td>
<td>letters to her fiancé/ husband F.S. Malan (7/12/1893, 16/11/1894, 15/2/1895, 15/3/1895, 10/5/1895, 29/3/1895, 14/5/1901, May 1901), to her sister-in-law Katie v. Huyssteen (22/12/1905)</td>
<td>3800</td>
<td>CA-A 583-8/ SAL-MSC 853-4-1</td>
</tr>
<tr>
<td>Brümmer, Johann J.</td>
<td>letter to his sister Johanna Brümmer (12/7/1900)</td>
<td>315</td>
<td>CA-A 583-85</td>
</tr>
<tr>
<td>Brümmer, Nikolaas J.</td>
<td>letters to F.S. Malan (12/11/1894, 8/4/1895)</td>
<td>572</td>
<td>CA-A 583-2</td>
</tr>
<tr>
<td>Burger, Jacobus A.</td>
<td>letter to Katie v. Huyssteen (1/1/1906)</td>
<td>326</td>
<td>SAL-MSC 853-4-3</td>
</tr>
<tr>
<td>Coopman, Cornelius</td>
<td>letter to the Zendings-Commissie (18/9/1909)</td>
<td>760</td>
<td>SK- S 5-2-57-1</td>
</tr>
<tr>
<td>Cronje, Daniel C.</td>
<td>letter to Katie v. Huyssteen (7/6/1906)</td>
<td>190</td>
<td>SAL-MSC 853-4-4</td>
</tr>
<tr>
<td>Cronje, Hendrik S.</td>
<td>letter to Katie v. Huyssteen (15/6/1906)</td>
<td>244</td>
<td>SAL-MSC 853-1-5</td>
</tr>
<tr>
<td>Cronje, Maria</td>
<td>letters to Katie v. Huyssteen (27/7/1906, 30/7/1906)</td>
<td>641</td>
<td>SAL-MSC 853-4-1/ MSB 853-1-5</td>
</tr>
<tr>
<td>Cronje, Wessels C.</td>
<td>letter to Daniel Malan (28/7/1905)</td>
<td>714</td>
<td>SAL-MSC 852-1-13</td>
</tr>
<tr>
<td>Dampies, Piet</td>
<td>letter to the Zendings-Commissie (30/8/1891)</td>
<td>337</td>
<td>SK- S 5-21-26</td>
</tr>
<tr>
<td>Name</td>
<td>Description of the Documents</td>
<td>No. of Words</td>
<td>Ref.</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>------------------------------------------------------------------------------------------------</td>
<td>--------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>Daniels, Sophie</td>
<td>letter to P.O. Henning (23/9/1900)</td>
<td>266</td>
<td>SAL-MSC 1-29-1</td>
</tr>
<tr>
<td>DeVilliers, J.DeW.</td>
<td>letter to C.P. Hoogenhout (11/3/1894)</td>
<td>182</td>
<td>SAL-MSC 7-8-4</td>
</tr>
<tr>
<td>DeVilliers, J.H.</td>
<td>letter to S.J. Du Toit (22/4/1891)</td>
<td>274</td>
<td>SAL-MSC 1-34-9</td>
</tr>
<tr>
<td>DeVilliers, Jan J.</td>
<td>letter to D.J. Du Toit (4/12/1882)</td>
<td>194</td>
<td>SAL-MSB 892-1-4</td>
</tr>
<tr>
<td>DeVos, Helena Cath.</td>
<td>letter to her niece S. Hendrikz (23/8/1899)</td>
<td>334</td>
<td></td>
</tr>
<tr>
<td>DeVries, Anna</td>
<td>letters to Katie vHuysteen (13/6/1906), to C.P. Hoogenhout (6/6/1899, 3/12/1901, 11/6/1906, 10/9/1906)</td>
<td>1224</td>
<td>SAL-MSB 853-1-5/MSC 1-24-3</td>
</tr>
<tr>
<td>DeWaal, Beatrice</td>
<td>letter to C.P. Hoogenhout (21/12/1908)</td>
<td>159</td>
<td>SAL-MSC 1-29-5</td>
</tr>
<tr>
<td>DeWaal, Jan H.H.</td>
<td>letters to F.S. Malan (28/10/1901, 8/1/1902), to C.P. Hoogenhout (3/6/1910)</td>
<td>1019</td>
<td>CA-A 583-2/SAL-MSC 1-36-4</td>
</tr>
<tr>
<td>DeWet, D.C.</td>
<td>letter to E. Esselen (24/11/1888)</td>
<td>970</td>
<td>UCT-BC 94-D3.20</td>
</tr>
<tr>
<td>DeWet, Nicolaas J.</td>
<td>letters to F.S. Malan (1894, 21/5/1895, June 1895)</td>
<td>1026</td>
<td>CA-A 583-2</td>
</tr>
<tr>
<td>Dippenaar, Johannes</td>
<td>letter to S. Hendrikz (22/12/1900)</td>
<td>365</td>
<td>SAL-MSB 892-1-3</td>
</tr>
<tr>
<td>Du Preez, Johannes G.</td>
<td>letter to his brother N.J. Du Preez (11/8/1894)</td>
<td>354</td>
<td>SAL-MSB 162-1</td>
</tr>
<tr>
<td>Du Preez, P.E.</td>
<td>letter to his brother N.J. Du Preez (24/8/1891)</td>
<td>144</td>
<td>SAL-MSB 162-1</td>
</tr>
<tr>
<td>Eksteen, Elizabeth J.H.</td>
<td>letters to her daughter Rachel (31/8/1919, 1/1/1922)</td>
<td>1565</td>
<td>SAL-MSC 33-3-1/MSC 33-5-1</td>
</tr>
<tr>
<td>Eksteen, P.J.</td>
<td>letter to P.O. Henning (15/3/1899)</td>
<td>324</td>
<td></td>
</tr>
<tr>
<td>Es(s)op, Napie</td>
<td>letters to C.P. Hoogenhout (23/4/1919, 5/7/1919)</td>
<td>191</td>
<td>G-brieue onderwyser aan Superintendent 1850-1900</td>
</tr>
<tr>
<td>Goosen, Jan A.B.</td>
<td>letters to Imker Hoogenhout (10/10/1908, 25/4/1909, 2/8/1909)</td>
<td>920</td>
<td>SAL-MSC 1-17-5/MSC 1-17-4</td>
</tr>
<tr>
<td>Goosen, Pieter A. sen.</td>
<td>letters to Imker Hoogenhout (27/7/1909), to C.P. Hoogenhout (2/2/1912)</td>
<td>400</td>
<td>SAL-MSC 1-17-4/MSC 1-27-4</td>
</tr>
<tr>
<td>Goosen, Sebastina</td>
<td>letter to her parents-in-law (17/6/1905, 7/4/1910, 19/4/1910), to her husband (20/3/1912)</td>
<td>1357</td>
<td>SAL-MSC1-10/MSC1-14-2/MSC1-14-5/MSC1-17-1</td>
</tr>
<tr>
<td>Goosen, S. J.F. sen</td>
<td>letter to C.P. Hoogenhout (1913)</td>
<td>202</td>
<td>SAL-MSC 1-22-3</td>
</tr>
<tr>
<td>Name</td>
<td>Description of the Documents</td>
<td>No. of Words</td>
<td>Ref.</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>------------------------------------------------------------------------------------------------</td>
<td>--------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Hatting, Benjamin</td>
<td>letter to the Sendings-Commissie (18/7/1886)</td>
<td>587</td>
<td>SK- 5-2-126</td>
</tr>
<tr>
<td>Henderson, James</td>
<td>letter to Marie Koopmans-DeWet (c. 1900)</td>
<td>257</td>
<td>CA- A 583-89</td>
</tr>
<tr>
<td>Hendrikz, Sophie</td>
<td>letter to her cousin 'Johannes' (c. 1900)</td>
<td>283</td>
<td>SAL- MSB 892-1-3</td>
</tr>
<tr>
<td>Heroldt, Michiel J.</td>
<td>letter to his grandchild 'Annie' (28/9/1891) and life story (circa 1900)</td>
<td>4026</td>
<td>SAL- MSB 229-1/ MSB 229-2</td>
</tr>
<tr>
<td>Hillebrand, Martha C.</td>
<td>letter to her grandchild Katie vHuysesteeen (13/10/1903)</td>
<td>591</td>
<td>SAL- MSB 853-1-2</td>
</tr>
<tr>
<td>Hiela, Johannes</td>
<td>letters to the Moravians in Genadendal (22/8/1906, 20/9/1906)</td>
<td>530</td>
<td>G - Buchkasse Briefe 1905-1907</td>
</tr>
<tr>
<td>Hoogenhout, Hendrina</td>
<td>letters to her-parents-in-law (22/1/1911, 19/1/1911, 14/2/1912, 13/5/1919)</td>
<td>965</td>
<td>SAL- MSC 1-18-6/MSC1-27-2/MSC1-18-4</td>
</tr>
<tr>
<td>Hoogenhout, Hester</td>
<td>letters to her-parents-in-law (13/2/1906, 8/11/1906, 18/11/1906, 21/2/1907, 10/3/1908, 17/5/1908, 10/11/1908)</td>
<td>1641</td>
<td>SAL- MSC 1-5-2/ MSC 1-5-4/ MSC 1-5-1</td>
</tr>
<tr>
<td>Hoogenhout, Johanna</td>
<td>letters to her parents (5/7/1893, 6/2/1896, 2/1/1897, 26/5/1899, 29/4/1902, 26/10/1902, 4/1/1902, 11/6/1905)</td>
<td>1651</td>
<td>SAL- MSC 1-10-2</td>
</tr>
<tr>
<td>Hoogenhout, Nicolaas M.</td>
<td>letters to his parents (10/3/1897, 29/4/1897, 14/4/1905, 20/5/1906, 24/6/1910)</td>
<td>2009</td>
<td>SAL- MSC 1-5-2/ MSC 1-5-3</td>
</tr>
<tr>
<td>Hoogenhout, Petrus I.</td>
<td>letters to his parents (24/5/1902, 8/5/1908, 23/5/1908, 25/5/1908), to his brother Imker (23/2/1896)</td>
<td>783</td>
<td>SAL- MSC 1-17-1/ MSC 1-17-5/ MSC 1-14-3</td>
</tr>
<tr>
<td>January, Wilhelmina</td>
<td>letter to P.O. Henning (30/3/1897)</td>
<td>553</td>
<td>G-brieue onderwysers aan Superintendent1850-1900</td>
</tr>
<tr>
<td>Koch J.A.</td>
<td>letter to Katie vHuysesteeen (27/7/1909)</td>
<td>355</td>
<td>SAL- MSE 853-1-8</td>
</tr>
<tr>
<td>Kotze, T.N.</td>
<td>letter to F.S. Malan (1/1/1904)</td>
<td>398</td>
<td>CA- A 583-4</td>
</tr>
<tr>
<td>Krige, Leopold J.</td>
<td>letter to his sister Isie Smuts (7/3/1907)</td>
<td>260</td>
<td>UCT- BCZA- 77/107/14</td>
</tr>
<tr>
<td>Kuhn, Pieter G.</td>
<td>letter to F.S. Malan (16/5/1900)</td>
<td>175</td>
<td>CA- A 583-4</td>
</tr>
<tr>
<td>Name</td>
<td>Description of the Documents</td>
<td>No. of Words</td>
<td>Ref.</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------------------------------------------------------------------------------------</td>
<td>--------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>LeRoux, D.J.</td>
<td>letters to C.P. Hooghenhou (15/11/1886, 17/11/1886)</td>
<td>253</td>
<td>SAL- MSC 1-41-1</td>
</tr>
<tr>
<td>LeRoux, Petrus J.</td>
<td>letters to his step father C.P. Hooghenhou (22/3/1905, 1/5/1908, 18/2/1908, 6/6/1908, 16/8/1909, 6/4/1912, 16/11/1913, three letters without date)</td>
<td>2035</td>
<td>SAL- - MSC 1-22-1/ MSC 1-29-1</td>
</tr>
<tr>
<td>Louw, Andries A. jun.</td>
<td>letter to his father A.A. Louw, sen. (28/2/1887)</td>
<td>720</td>
<td>NGK- P 52-1-2</td>
</tr>
<tr>
<td>Louw, Francina S.</td>
<td>letter to her father in law A.A. Louw, sen. (14/3/1899)</td>
<td>591</td>
<td>NGK- P 52-1-2</td>
</tr>
<tr>
<td>Louw, Maria J.</td>
<td>letter to her father, A.A. Louw, sen. (27/5/1901)</td>
<td>297</td>
<td>SAL- MSB 57</td>
</tr>
<tr>
<td>Louw, Rijkie Hester</td>
<td>letter to Katie vHuyssteen (4/8/1909)</td>
<td>337</td>
<td>SAL- MSB 853-1-8</td>
</tr>
<tr>
<td>Louwrens, Anna S.</td>
<td>letter to Katie vHuyssteen (16/3/1905)</td>
<td>150</td>
<td>SAL- MSB 853-1-3</td>
</tr>
<tr>
<td>Lourens, Joh.F.</td>
<td>letter to Daniel Malan (28/12/1905)</td>
<td>189</td>
<td>SAL- MSB 853-5-1</td>
</tr>
<tr>
<td>Malan, Daniel G.</td>
<td>letters to his daughter-in-law Katie van Huyssteen (29/5/1906, 22/6/1906, 14/12/1906, 25/8/1907)</td>
<td>684</td>
<td>SAL- MSB 853-1-5/ MSB 853-1-6</td>
</tr>
<tr>
<td>Malan, Elizabeth J.</td>
<td>letters to her son F.S. Malan (9/5/1893, circa 1900), to her daughter-in-law Johanna Brümmer (16/12/1893), to her daughter-in-law Katie vHuyssteen (17/12/1905, 28/5/1906, 22/6/1906, 20/8/1907, no date), to her grandchildren Vrede, Richie and David Malan (27/5/1909)</td>
<td>1749</td>
<td>CA- A 583-7/ A 583-85</td>
</tr>
<tr>
<td>Malan, Francois S.</td>
<td>extracts from diaries (1892-1895), letter to W. Viljoen (transcript of the letter in his diary 4/10/1894), letters to his fiancee/ wife Johanna Brümmer (10/4/1892, 1/6/1893, 31/10/1894, 28/3/1895, 2/11/1894, 1/6/1893, 7/4/1895, 2/4/1896, 7/6/1896, 27/12/1906), letter to his sister-in-law Katie vHuyssteen (20/1/1906)</td>
<td>3906</td>
<td>CA - A 583-16/ A 583-17/ A 583-18/ A-583-83/ A 583-84/SA- MSB 853-4-1</td>
</tr>
<tr>
<td>Malan, Jacques</td>
<td>postcard to his cousin Vrede Malan (circa 1910)</td>
<td>328</td>
<td>SAL- MSB 853 (reprinted in Quart Bull S Afr Libr 49 (3) 1995: 128-9)</td>
</tr>
<tr>
<td>Malan, Lelon</td>
<td>letter to C.P. Hooghenhou (31/10/1917)</td>
<td>465</td>
<td>SAL- MSC 1-29-1</td>
</tr>
<tr>
<td>Malan, Maria C.</td>
<td>letters to her brother F.S. Malan (14/5/1892, 23/4/1893), sister-in-law Katie vHuyssteen (22/12/1907, 29/5/1906), to her nephews Vrede, Rechter and David (4/11/1906)</td>
<td>1759</td>
<td>CA- A 583-7</td>
</tr>
<tr>
<td>Marais, F.P.</td>
<td>to C.P. Hooghenhou (23/1/1895)</td>
<td>236</td>
<td>NGK- S 5-2-13</td>
</tr>
<tr>
<td>Marais, Helena F.</td>
<td>letters to her brother-in-law C.P. Hooghenhou and his wife (her sister) (23/1/1896, 10/2/1896)</td>
<td>564</td>
<td>SAL - MSC 1-22-3/ MSC 1-22-4</td>
</tr>
<tr>
<td>Name</td>
<td>Description of the Documents</td>
<td>No. of Words</td>
<td>Ref.</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>------------------------------------------------------------------------------------------------</td>
<td>--------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>Marais, Rachel</td>
<td>letter to C.P. Hoogenhout 8/5/1909)</td>
<td>383</td>
<td>SAL- MSC 1-27-3</td>
</tr>
<tr>
<td>Marais, Stephanus J.</td>
<td>letter to C. Rayner (5/6/1901)</td>
<td>283</td>
<td>SAL- MSB 744-1-39</td>
</tr>
<tr>
<td>Marren, 'Maggie'</td>
<td>letters to C.P. Hoogenhout (24/3/1915, 13/6/1915)</td>
<td>537</td>
<td>SAL- MSC 1-27-5</td>
</tr>
<tr>
<td>Matolana, Mozes</td>
<td>letter to P.O. Henning (5/7/1898)</td>
<td>311</td>
<td>G-brieue onderwyser aan superintendent 1850-1900</td>
</tr>
<tr>
<td>Mlimkulu, Mozes</td>
<td>letter to P.O. Henning (1/12/1896)</td>
<td>575</td>
<td>G-brieue onderwyser aan superintendent 1850-1900</td>
</tr>
<tr>
<td>Muller, Hermanus Steyn</td>
<td>letter to Katie vHuysteen and David Malan (18/5/1905)</td>
<td>431</td>
<td>SAL- MSB 853-1-3</td>
</tr>
<tr>
<td>November, Jacobus</td>
<td>letter to C.P. Hoogenhout (13/2/1912)</td>
<td>100</td>
<td>SAL- MSC 1-26-4</td>
</tr>
<tr>
<td>Odendaal, Johanna</td>
<td>letters to Katie vHuysteen (2/2/1906)</td>
<td>175</td>
<td>SAL- MSB 853-4-1</td>
</tr>
<tr>
<td>Oosthuizen, M. A. M.</td>
<td>letters to Katie vHuysteen (3/1/1906, 27/12/1905)</td>
<td>1007</td>
<td>SAL- MSB 853-4-1/ 853-1-3</td>
</tr>
<tr>
<td>Pauw, Anna</td>
<td>letter to C.P. Hoogenhout (17/6/1917)</td>
<td>299</td>
<td>SAL- MSC 1-29-5</td>
</tr>
<tr>
<td>Pienaar, Hendrik H.</td>
<td>letter to David Malan (17/5/1905)</td>
<td>347</td>
<td>SAL- MSC 583-1-3</td>
</tr>
<tr>
<td>Roos, 'Lenie'</td>
<td>letter to Marie Koopmans-DeWet (29/1/1901)</td>
<td>214</td>
<td>CA- A 583-89</td>
</tr>
<tr>
<td>Roussouw, Aletta M.</td>
<td>letter to her brother-law C.P. Hoogenhout and his wife (her sister) (4/4/1904)</td>
<td>204</td>
<td>SAL- MSC 1-29-1</td>
</tr>
<tr>
<td>Roux, Katie</td>
<td>letter to J.W.C. Hofmeyer (27/12/1919)</td>
<td>466</td>
<td>NGK- S 5-2-18</td>
</tr>
<tr>
<td>Schabort, Elizabeth A.S.</td>
<td>description of her experiences before and during detention at Merebank Camp, Durban</td>
<td>4569</td>
<td>SAL- MSB 437-1-1</td>
</tr>
<tr>
<td>Schrade, Hester</td>
<td>letter to J.W.C. Hofmeyer (12/2/1920)</td>
<td>164</td>
<td>NGK- S 5-2-56-1</td>
</tr>
<tr>
<td>Small, Nicolaas</td>
<td>letters to C.P. Hoogenhout (1/7/1898), to Wyanda Hoogenhout (27/6/1898)</td>
<td>534</td>
<td>SAL- MSC 1-26-4/ MSC 1-21-6</td>
</tr>
<tr>
<td>Smith, Edmund T.</td>
<td>letters to his father-in-law B.J. Britimmer (circa 1900, 27/10/1904)</td>
<td>790</td>
<td>CA- A 583-3/ A 583-85</td>
</tr>
<tr>
<td>Smuts, J. A. ('Koos')</td>
<td>letter to his brother J.C. Smuts (11/5/1902)</td>
<td>289</td>
<td>UCT- BCZA 77-104-102</td>
</tr>
<tr>
<td>Smuts, J.C.</td>
<td>letter to his brother J.A. Smuts (4/1/1902)</td>
<td>620</td>
<td>UCT- BCZA-77-104-108</td>
</tr>
<tr>
<td>Smuts, Syabella M. ('Isie')</td>
<td>letters to R.I. Steyn (9/9/1912, 13/11/1912)</td>
<td>563</td>
<td>SAL- MSC 33-17-2</td>
</tr>
<tr>
<td>Stanford, Cornelia</td>
<td>letters to Katie vHuyssteen (6/8/1905, 9/6/1906)</td>
<td>1090</td>
<td>SAL- MSB 853-1-3/ MSB 853-4-1</td>
</tr>
<tr>
<td>Steyn, Anna</td>
<td>letter to Katie vHuyssteen (16/3/1906)</td>
<td>445</td>
<td>SAL- MSB 853-4-3</td>
</tr>
<tr>
<td>Steyn, Daniel</td>
<td>letter to Katie vHuyssteen (2/6/1906)</td>
<td>208</td>
<td>SAL- MSB 853-1-5</td>
</tr>
<tr>
<td>Steyn, Elizabeth G.</td>
<td>letters to Katie vHuyssteen (2/6/1906, 16/3/1905)</td>
<td>1100</td>
<td>SAL- MSB 853-1-5/ MSB 853-1-3</td>
</tr>
<tr>
<td>Name</td>
<td>Description of the Documents</td>
<td>No. of Words</td>
<td>Ref.</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------------------------------------------------------------------------------------------</td>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Steyn, Rachel I.</td>
<td>letters to her (future) daughter-in-law Rachel Eksteen (4/6/1919, 19/10/1919, 4/12/1919)</td>
<td>1613</td>
<td>SAL- MSC 33-17-1</td>
</tr>
<tr>
<td>Uys, Susan</td>
<td>letter to Katie vHuyssteen (8/1/1906)</td>
<td>275</td>
<td>SAL- MSB 853-4-3</td>
</tr>
<tr>
<td>vd Merwe, C.J.</td>
<td>letters to D.S. DeVilliers (26/11/1900, 30/11/1900, 3/12/1900, 4/12/1900)</td>
<td>667</td>
<td>SAL- MSB 510-1-1</td>
</tr>
<tr>
<td>vHuyssteen, A.J.Z. ('Janie')</td>
<td>letter to her sister Katie vHuyssteen (31/12/1907)</td>
<td>400</td>
<td>SAL- MSB 853-1-6</td>
</tr>
<tr>
<td>vHuyssteen, Anna Chr.</td>
<td>letter to her sister-in-law Katie vHuyssteen (23/7/1907)</td>
<td>263</td>
<td>SAL- MSB 583-1-6</td>
</tr>
<tr>
<td>vHuyssteen, Dan. P.S.</td>
<td>letters to his sister Katie vHuyssteen (17/1/1906, 16/6/1906, 30/6/1909)</td>
<td>378</td>
<td>SAL- MSB 853-1-5/MSB 853-1-5</td>
</tr>
<tr>
<td>vHuyssteen, Eliza J.</td>
<td>letters to her sister Katie vHuyssteen (2/3/1907, 16/4/1908)</td>
<td>1356</td>
<td>SAL- MSB 853-1-6</td>
</tr>
<tr>
<td>vHuyssteen, E.C.P.</td>
<td>letters to her daughter Katie vHuyssteen (10 letter 1905/06), to her daughter J.H. vHuyssteen (1906)</td>
<td>4400</td>
<td>SAL- MSB 853-1-5/MSB 853-1-10/MSB 853-1-3</td>
</tr>
<tr>
<td>vHuyssteen, Hendrik</td>
<td>letter to his daughter Katie vHuyssteen (9/1/1906, 3/10/1908), to his grandson David (14/10/1909)</td>
<td>787</td>
<td>SAL- MSB 853-1-5/MSB 853-1-3/MSB 853-1-8</td>
</tr>
<tr>
<td>vHuyssteen, J.P.J.W.</td>
<td>letters to his sister Katie vHuyssteen (20/5/1907, 27/7/1909)</td>
<td>1221</td>
<td>SAL- MSB 853-1-6/MSB 853-1-8</td>
</tr>
<tr>
<td>vHuyssteen, Jacomina H.</td>
<td>letters to her sister Katie vHuyssteen (14/4/1908, 5/11/1909)</td>
<td>1364</td>
<td>SAL- MSB 583-1-6/MSB 853-1-8</td>
</tr>
<tr>
<td>vHuyssteen, M.C. ('Katie')</td>
<td>letters to her parents (21/4/1906, 6/8/1906, May 1907)</td>
<td>1375</td>
<td>SAL-MSB 853-1-1</td>
</tr>
<tr>
<td>vHuyssteen, Petrus H.H.</td>
<td>letters to his sister Katie vHuyssteen (21/2/1906, Dec. 1905)</td>
<td>727</td>
<td>SAL- MSB 853-4-3/MSB 853-4-1</td>
</tr>
<tr>
<td>vHuyssteen, Matthys P.Z.</td>
<td>letter to his sister Katie vHuyssteen (10/1/1905)</td>
<td>303</td>
<td>SAL- MSB 853-4-1</td>
</tr>
<tr>
<td>vOosthuizen, Lettie</td>
<td>letter to Katie vHuyssteen (5/3/1907)</td>
<td>286</td>
<td>SAL- MSB 1-6</td>
</tr>
<tr>
<td>vZyl, Susan</td>
<td>letter to Katie vHuyssteen (24/3/1905)</td>
<td>361</td>
<td>SAL- MSB 853-1-3</td>
</tr>
<tr>
<td>Vermaak, Hester S.</td>
<td>letters to Katie vHuyssteen (Nov. 1905, 27/1/1906, 19/2/1906, 13/6/1907, 22/6/1907)</td>
<td>2223</td>
<td>SAL- MSB 853-4-1/MSB 853-4-3/MSB 853-1-6/</td>
</tr>
<tr>
<td>'Hester'</td>
<td>to 'Maria' (7/11/1900)</td>
<td>398</td>
<td>UCT: B28 F19</td>
</tr>
<tr>
<td>'Gay References'</td>
<td>to Catharina Maria Hoogenhout (circa 1900)</td>
<td>173</td>
<td>SAL- MSC 1-26-4</td>
</tr>
<tr>
<td>'Auntie'</td>
<td>letter to her niece (27/4/1904)</td>
<td>376</td>
<td>SAL- MSB 744-1-43</td>
</tr>
</tbody>
</table>

325
## B 3: THE DATA SET
(number of 'Afrikaans' variants/ total number of loci; exception VIR, i.e. number of VIR/ total number of words)

<table>
<thead>
<tr>
<th>name</th>
<th>Gender</th>
<th>Verb(main)</th>
<th>Infinitive</th>
<th>PP - t</th>
<th>PP - en</th>
<th>Tense</th>
<th>Pronoun*</th>
<th>Om. te</th>
<th>Nie-2</th>
<th>Vir</th>
</tr>
</thead>
<tbody>
<tr>
<td>April, Aletta</td>
<td>9/15</td>
<td>2/4</td>
<td>8/19</td>
<td>0/9</td>
<td>0/9</td>
<td>14/19</td>
<td>1/30</td>
<td>3/6</td>
<td>0/7</td>
<td>0/-</td>
</tr>
<tr>
<td>April, Anton</td>
<td>2/2</td>
<td>0/1</td>
<td>1/3</td>
<td>1/3</td>
<td>1/2</td>
<td>4/4</td>
<td>0/16</td>
<td>1/1</td>
<td>2/2</td>
<td>0/-</td>
</tr>
<tr>
<td>Baalke, Fredrik</td>
<td>0/1</td>
<td>-</td>
<td>0/7</td>
<td>0/1</td>
<td>0/2</td>
<td>3/4</td>
<td>0/7</td>
<td>3/4</td>
<td>0/3</td>
<td>0/-</td>
</tr>
<tr>
<td>Baalke, Rudolf</td>
<td>1/4</td>
<td>1/11</td>
<td>0/33</td>
<td>0/6</td>
<td>0/1</td>
<td>4/13</td>
<td>0/27</td>
<td>3/11</td>
<td>0/7</td>
<td>1/639</td>
</tr>
<tr>
<td>Baartman, Susie [1]</td>
<td>0/2</td>
<td>0/2</td>
<td>0/4</td>
<td>0/4</td>
<td>0/2</td>
<td>4/5</td>
<td>0/11</td>
<td>2/2</td>
<td>0/2</td>
<td>0/-</td>
</tr>
<tr>
<td>Baartman, Susie [2]</td>
<td>1/1</td>
<td>7/7</td>
<td>12/12</td>
<td>2/2</td>
<td>3/3</td>
<td>7/7</td>
<td>16/20</td>
<td>3/3</td>
<td>0/2</td>
<td>2/402</td>
</tr>
<tr>
<td>Basson, [Mary]</td>
<td>2/9</td>
<td>8/23</td>
<td>2/67</td>
<td>0/6</td>
<td>18/24</td>
<td>1/69</td>
<td>4/15</td>
<td>0/31</td>
<td>2/1644</td>
<td></td>
</tr>
<tr>
<td>Beyers, C. E. [1]</td>
<td>4/14</td>
<td>0/18</td>
<td>1/22</td>
<td>2/9</td>
<td>8/17</td>
<td>0/37</td>
<td>8/10</td>
<td>0/19</td>
<td>0/-</td>
<td>0/-</td>
</tr>
<tr>
<td>Blake, Ebenzer J.</td>
<td>3/4</td>
<td>-</td>
<td>9/14</td>
<td>-</td>
<td>0/3</td>
<td>2/2</td>
<td>0/10</td>
<td>1/3</td>
<td>0/2</td>
<td>0/-</td>
</tr>
<tr>
<td>Bomhuis, Jan</td>
<td>-</td>
<td>0/1</td>
<td>0/4</td>
<td>2/2</td>
<td>0/1</td>
<td>2/2</td>
<td>0/5</td>
<td>2/2</td>
<td>0/2</td>
<td>0/-</td>
</tr>
<tr>
<td>Bosman, Louisa</td>
<td>1/1</td>
<td>1/1</td>
<td>6/6</td>
<td>1/2</td>
<td>-</td>
<td>1/1</td>
<td>9/11</td>
<td>4/4</td>
<td>0/1</td>
<td>2/161</td>
</tr>
<tr>
<td>Botha, Louis</td>
<td>7/15</td>
<td>13/38</td>
<td>16/89</td>
<td>1/11</td>
<td>1/10</td>
<td>16/23</td>
<td>13/98</td>
<td>15/27</td>
<td>0/21</td>
<td>0/-</td>
</tr>
<tr>
<td>Breuris, Frans M.C.</td>
<td>1/1</td>
<td>3/14</td>
<td>6/37</td>
<td>1/5</td>
<td>0/4</td>
<td>7/10</td>
<td>9/42</td>
<td>1/16</td>
<td>0/14</td>
<td>3/910</td>
</tr>
<tr>
<td>Brümmer, Barend J.</td>
<td>12/22</td>
<td>23/38</td>
<td>46/91</td>
<td>7/18</td>
<td>10/29</td>
<td>41/41</td>
<td>7/120</td>
<td>8/20</td>
<td>0/26</td>
<td>5/3122</td>
</tr>
<tr>
<td>Brümmer Johanna [1]</td>
<td>2/5</td>
<td>10/21</td>
<td>1/30</td>
<td>0/6</td>
<td>0/12</td>
<td>13/19</td>
<td>0/36</td>
<td>0/7</td>
<td>0/14</td>
<td>1/960</td>
</tr>
<tr>
<td>Brümmer, Nikolaas J.</td>
<td>0/1</td>
<td>1/3</td>
<td>2/12</td>
<td>0/4</td>
<td>0/6</td>
<td>6/9</td>
<td>0/14</td>
<td>5/6</td>
<td>0/4</td>
<td>0/-</td>
</tr>
<tr>
<td>Burger, Jacobus A.</td>
<td>1/3</td>
<td>4/6</td>
<td>1/8</td>
<td>1/1</td>
<td>0/6</td>
<td>4/5</td>
<td>1/23</td>
<td>2/3</td>
<td>0/5</td>
<td>0/-</td>
</tr>
<tr>
<td>Cillie, Francois S.</td>
<td>5/6</td>
<td>5/9</td>
<td>10/15</td>
<td>7/10</td>
<td>3/12</td>
<td>19/19</td>
<td>7/27</td>
<td>2/3</td>
<td>0/9</td>
<td>0/-</td>
</tr>
<tr>
<td>Coopman, Cornelius</td>
<td>1/1</td>
<td>6/8</td>
<td>14/18</td>
<td>11/13</td>
<td>10/14</td>
<td>1/2</td>
<td>26/37</td>
<td>7/7</td>
<td>13/14</td>
<td>2/760</td>
</tr>
<tr>
<td>Cronje, Daniel C.</td>
<td>1/1</td>
<td>3/3</td>
<td>5/6</td>
<td>0/1</td>
<td>1/1</td>
<td>2/2</td>
<td>4/13</td>
<td>3/3</td>
<td>1/1</td>
<td>0/-</td>
</tr>
<tr>
<td>Cronje, Hendrik S.</td>
<td>1/3</td>
<td>6/8</td>
<td>1/8</td>
<td>2/2</td>
<td>0/2</td>
<td>6/6</td>
<td>1/21</td>
<td>1/2</td>
<td>0/6</td>
<td>0/-</td>
</tr>
<tr>
<td>Cronje, Maria</td>
<td>0/4</td>
<td>2/12</td>
<td>2/35</td>
<td>1/6</td>
<td>0/9</td>
<td>9/9</td>
<td>0/40</td>
<td>6/16</td>
<td>0/10</td>
<td>0/-</td>
</tr>
<tr>
<td>Cronje, Wessels C.</td>
<td>2/5</td>
<td>21/21</td>
<td>9/18</td>
<td>5/9</td>
<td>0/6</td>
<td>8/12</td>
<td>0/16</td>
<td>5/8</td>
<td>0/3</td>
<td>0/-</td>
</tr>
<tr>
<td>Dampies, Piet</td>
<td>2/2</td>
<td>3/4</td>
<td>3/9</td>
<td>1/4</td>
<td>4/4</td>
<td>8/8</td>
<td>6/16</td>
<td>3/4</td>
<td>1/1</td>
<td>0/-</td>
</tr>
<tr>
<td>Name</td>
<td>Gender</td>
<td>Verb (main)</td>
<td>Infinitive</td>
<td>PP - t</td>
<td>PP - en</td>
<td>Tense</td>
<td>Pronoun*</td>
<td>Om...te</td>
<td>Nie-2</td>
<td>Vir</td>
</tr>
<tr>
<td>----------------</td>
<td>--------</td>
<td>------------</td>
<td>------------</td>
<td>--------</td>
<td>--------</td>
<td>-------</td>
<td>----------</td>
<td>--------</td>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>Daniels, Sophie</td>
<td>-</td>
<td>1/2</td>
<td>1/8</td>
<td>0/3</td>
<td>0/6</td>
<td>7/7</td>
<td>1/17</td>
<td>0/4</td>
<td>0/4</td>
<td>1/266</td>
</tr>
<tr>
<td>De Villiers, J. De W.</td>
<td>0/3</td>
<td>0/3</td>
<td>1/9</td>
<td>1/1</td>
<td>0/3</td>
<td>2/2</td>
<td>1/11</td>
<td>0/5</td>
<td>0/2</td>
<td>0/-</td>
</tr>
<tr>
<td>De Villiers, J.H.</td>
<td>3/4</td>
<td>4/4</td>
<td>1/11</td>
<td>5/6</td>
<td>0/2</td>
<td>5/6</td>
<td>1/8</td>
<td>3/3</td>
<td>4/4</td>
<td>0/-</td>
</tr>
<tr>
<td>De Villiers, Jan J.</td>
<td>1/2</td>
<td>-</td>
<td>0/5</td>
<td>0/4</td>
<td>0/2</td>
<td>1/6</td>
<td>0/6</td>
<td>1/1</td>
<td>0/1</td>
<td>0/-</td>
</tr>
<tr>
<td>De Vos, Helena Cath.</td>
<td>4/4</td>
<td>3/3</td>
<td>1/15</td>
<td>3/3</td>
<td>1/1</td>
<td>2/2</td>
<td>0/11</td>
<td>4/6</td>
<td>0/6</td>
<td>0/-</td>
</tr>
<tr>
<td>DeVries, Anna</td>
<td>1/10</td>
<td>2/31</td>
<td>0/39</td>
<td>0/8</td>
<td>0/9</td>
<td>11/23</td>
<td>1/87</td>
<td>2/9</td>
<td>0/19</td>
<td>0/-</td>
</tr>
<tr>
<td>De Waal, Beatrice</td>
<td>-</td>
<td>2/2</td>
<td>2/10</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>De Waal, Jan H.H. [1]</td>
<td>1/12</td>
<td>0/12</td>
<td>0/29</td>
<td>0/9</td>
<td>0/12</td>
<td>11/20</td>
<td>1/15</td>
<td>5/12</td>
<td>0/14</td>
<td>0/-</td>
</tr>
<tr>
<td>De Waal, Jan H.H. [2]</td>
<td>2/2</td>
<td>1/1</td>
<td>2/2</td>
<td>2/2</td>
<td>1/1</td>
<td>3/3</td>
<td>3/3</td>
<td>3/3</td>
<td>3/3</td>
<td>0/-</td>
</tr>
<tr>
<td>De Wet, D.C.</td>
<td>9/9</td>
<td>13/14</td>
<td>29/40</td>
<td>11/14</td>
<td>8/11</td>
<td>28/29</td>
<td>34/45</td>
<td>7/10</td>
<td>7/8</td>
<td>2/970</td>
</tr>
<tr>
<td>Dippenaar, Johannes</td>
<td>0/2</td>
<td>1/1</td>
<td>0/9</td>
<td>1/2</td>
<td>0/6</td>
<td>5/9</td>
<td>3/22</td>
<td>4/6</td>
<td>0/2</td>
<td>0/-</td>
</tr>
<tr>
<td>Du Preez, Johannes G.</td>
<td>3/3</td>
<td>3/3</td>
<td>4/5</td>
<td>4/4</td>
<td>3/6</td>
<td>10/10</td>
<td>4/8</td>
<td>1/6</td>
<td>0/3</td>
<td>1/354</td>
</tr>
<tr>
<td>Du Preez, P.E.</td>
<td>-</td>
<td>1/1</td>
<td>5/5</td>
<td>2/2</td>
<td>-</td>
<td>2/2</td>
<td>2/6</td>
<td>0/1</td>
<td>0/5</td>
<td>1/144</td>
</tr>
<tr>
<td>Eksteen, Elizabeth J.H.</td>
<td>11/11</td>
<td>13/21</td>
<td>32/56</td>
<td>5/9</td>
<td>12/14</td>
<td>27/32</td>
<td>37/85</td>
<td>2/19</td>
<td>0/14</td>
<td>1/1565</td>
</tr>
<tr>
<td>Eksteen, P.J.</td>
<td>3/6</td>
<td>1/2</td>
<td>0/16</td>
<td>1/9</td>
<td>0/1</td>
<td>7/10</td>
<td>1/7</td>
<td>1/4</td>
<td>0/7</td>
<td>0/-</td>
</tr>
<tr>
<td>Es(s)op, Napie</td>
<td>1/1</td>
<td>1/3</td>
<td>3/4</td>
<td>3/5</td>
<td>2/2</td>
<td>1/1</td>
<td>4/12</td>
<td>0/2</td>
<td>0/4</td>
<td>0/-</td>
</tr>
<tr>
<td>Furter, Engela J.M.</td>
<td>6/7</td>
<td>10/12</td>
<td>15/20</td>
<td>5/6</td>
<td>3/7</td>
<td>10/11</td>
<td>9/32</td>
<td>2/4</td>
<td>0/16</td>
<td>1/722</td>
</tr>
<tr>
<td>Goosen, Gideon J.C.</td>
<td>19/20</td>
<td>41/42</td>
<td>78/82</td>
<td>22/23</td>
<td>6/11</td>
<td>30/30</td>
<td>95/96</td>
<td>20/25</td>
<td>23/25</td>
<td>122337</td>
</tr>
<tr>
<td>Goosen, Jan A.B.</td>
<td>7/7</td>
<td>17/17</td>
<td>26/28</td>
<td>4/4</td>
<td>5/6</td>
<td>13/13</td>
<td>42/44</td>
<td>2/4</td>
<td>2/10</td>
<td>7/920</td>
</tr>
<tr>
<td>Goosen, Pieter A. jun.[1]</td>
<td>0/8</td>
<td>0/11</td>
<td>0/31</td>
<td>0/13</td>
<td>0/11</td>
<td>18/22</td>
<td>1/47</td>
<td>0/10</td>
<td>0/21</td>
<td>0/-</td>
</tr>
<tr>
<td>Goosen, Pieter A. jun. [2]</td>
<td>2/2</td>
<td>5/5</td>
<td>7/8</td>
<td>1/1</td>
<td>1/1</td>
<td>3/3</td>
<td>9/9</td>
<td>1/2</td>
<td>0/1</td>
<td>0/-</td>
</tr>
<tr>
<td>Goosen, Pieter A. sen.</td>
<td>4/4</td>
<td>6/8</td>
<td>7/14</td>
<td>3/3</td>
<td>3/7</td>
<td>10/11</td>
<td>6/21</td>
<td>1/3</td>
<td>1/2</td>
<td>1/400</td>
</tr>
<tr>
<td>Goosen, Sebastina [1]</td>
<td>0/1</td>
<td>0/11</td>
<td>0/7</td>
<td>0/1</td>
<td>0/4</td>
<td>6/7</td>
<td>0/13</td>
<td>0/2</td>
<td>0/3</td>
<td>0/-</td>
</tr>
<tr>
<td>Goosen, S. J.F.sen</td>
<td>-</td>
<td>2/2</td>
<td>6/6</td>
<td>1/3</td>
<td>1/1</td>
<td>4/4</td>
<td>1/8</td>
<td>0/2</td>
<td>0/4</td>
<td>0/-</td>
</tr>
<tr>
<td>Grosskopf, J.F.W.</td>
<td>56/56</td>
<td>116/120</td>
<td>182/193</td>
<td>43/45</td>
<td>41/54</td>
<td>103/104</td>
<td>240/240</td>
<td>36/38</td>
<td>86/90</td>
<td>166362</td>
</tr>
<tr>
<td>Hauptfleisch, A.P. [1]</td>
<td>7/20</td>
<td>16/30</td>
<td>17/48</td>
<td>1/10</td>
<td>2/10</td>
<td>12/12</td>
<td>18/67</td>
<td>12/23</td>
<td>2/19</td>
<td>0/-</td>
</tr>
<tr>
<td>Name</td>
<td>Gender</td>
<td>Verb(main)</td>
<td>Infinitive</td>
<td>PP - t</td>
<td>PP - en</td>
<td>Tense</td>
<td>Pronoun*</td>
<td>Om...te</td>
<td>Nie-2</td>
<td>Vir</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>--------</td>
<td>------------</td>
<td>-----------</td>
<td>--------</td>
<td>---------</td>
<td>-------</td>
<td>----------</td>
<td>---------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>Hatting, Benjamin</td>
<td>7/10</td>
<td>13/13</td>
<td>5/11</td>
<td>5/6</td>
<td>2/7</td>
<td>6/6</td>
<td>2/22</td>
<td>2/3</td>
<td>0/8</td>
<td>0/-</td>
</tr>
<tr>
<td>Henderson, James</td>
<td>-</td>
<td>-</td>
<td>7/10</td>
<td>0/3</td>
<td>1/2</td>
<td>2/2</td>
<td>1/15</td>
<td>2/2</td>
<td>0/3</td>
<td>2/257</td>
</tr>
<tr>
<td>Hendrikz, Sophie</td>
<td>0/1</td>
<td>3/7</td>
<td>0/13</td>
<td>0/1</td>
<td>1/1</td>
<td>2/2</td>
<td>1/16</td>
<td>1/5</td>
<td>0/5</td>
<td>0/-</td>
</tr>
<tr>
<td>Heroldt, Michiel J.</td>
<td>1/74</td>
<td>2/18</td>
<td>0/126</td>
<td>0/38</td>
<td>0/42</td>
<td>40/141</td>
<td>1/171</td>
<td>3/5</td>
<td>0/48</td>
<td>0/-</td>
</tr>
<tr>
<td>Hillebrand, Martha C.</td>
<td>2/3</td>
<td>8/8</td>
<td>6/7</td>
<td>6/6</td>
<td>1/3</td>
<td>13/13</td>
<td>8/25</td>
<td>4/4</td>
<td>0/8</td>
<td>3/591</td>
</tr>
<tr>
<td>Hlela, Johannes</td>
<td>8/10</td>
<td>2/2</td>
<td>14/24</td>
<td>2/3</td>
<td>1/7</td>
<td>6/6</td>
<td>1/17</td>
<td>4/7</td>
<td>0/3</td>
<td>1/523</td>
</tr>
<tr>
<td>Hoogenhout, Dan. M. [1]</td>
<td>1/2</td>
<td>0/5</td>
<td>0/8</td>
<td>0/3</td>
<td>0/1</td>
<td>6/6</td>
<td>0/18</td>
<td>0/4</td>
<td>0/11</td>
<td>0/-</td>
</tr>
<tr>
<td>Hoogenhout, F. P. [1]</td>
<td>0/24</td>
<td>0/13</td>
<td>0/44</td>
<td>0/11</td>
<td>0/16</td>
<td>22/27</td>
<td>0/36</td>
<td>1/11</td>
<td>0/19</td>
<td>0/-</td>
</tr>
<tr>
<td>Hoogenhout, Hendr. [1]</td>
<td>3/8</td>
<td>0/6</td>
<td>1/18</td>
<td>0/4</td>
<td>0/8</td>
<td>12/24</td>
<td>0/42</td>
<td>3/7</td>
<td>0/6</td>
<td>0/-</td>
</tr>
<tr>
<td>Hoogenhout, Hendr. [2]</td>
<td>1/1</td>
<td>5/5</td>
<td>6/6</td>
<td>1/1</td>
<td>-</td>
<td>14/14</td>
<td>0/1</td>
<td>-</td>
<td>0/-</td>
<td>0/-</td>
</tr>
<tr>
<td>Hoogenhout, Hester</td>
<td>15/16</td>
<td>36/38</td>
<td>55/59</td>
<td>9/12</td>
<td>13/13</td>
<td>20/22</td>
<td>7/481</td>
<td>2/23</td>
<td>9/24</td>
<td>3/1761</td>
</tr>
<tr>
<td>Hoogenhout, Johanna[1]</td>
<td>2/16</td>
<td>0/16</td>
<td>0/33</td>
<td>0/2</td>
<td>1/17</td>
<td>20/27</td>
<td>1/36</td>
<td>0/29</td>
<td>0/-</td>
<td>0/-</td>
</tr>
<tr>
<td>Hoogenhout, Johanna[2]</td>
<td>2/2</td>
<td>6/11</td>
<td>7/12</td>
<td>1/1</td>
<td>2/3</td>
<td>5/5</td>
<td>6/14</td>
<td>0/8</td>
<td>1/9</td>
<td>1/6</td>
</tr>
<tr>
<td>Hoogenhout, Nicolaas M.</td>
<td>0/14</td>
<td>0/26</td>
<td>0/65</td>
<td>0/17</td>
<td>0/10</td>
<td>17/19</td>
<td>2/74</td>
<td>7/28</td>
<td>0/32</td>
<td>0/-</td>
</tr>
<tr>
<td>Hoogenhout, Petrus I.[1]</td>
<td>0/7</td>
<td>0/6</td>
<td>0/18</td>
<td>0/7</td>
<td>0/5</td>
<td>6/7</td>
<td>0/19</td>
<td>3/8</td>
<td>0/9</td>
<td>0/-</td>
</tr>
<tr>
<td>Hoogenhout, Petrus I.[2]</td>
<td>2/2</td>
<td>8/8</td>
<td>5/8</td>
<td>1/1</td>
<td>2/2</td>
<td>6/6</td>
<td>1/1</td>
<td>2/5</td>
<td>0/2</td>
<td>5/236</td>
</tr>
<tr>
<td>Hoogenhout, W. [1]</td>
<td>3/9</td>
<td>0/20</td>
<td>0/25</td>
<td>0/6</td>
<td>0/8</td>
<td>12/23</td>
<td>0/46</td>
<td>7/10</td>
<td>0/11</td>
<td>0/-</td>
</tr>
<tr>
<td>January, Wilhelmina</td>
<td>1/3</td>
<td>1/7</td>
<td>2/16</td>
<td>1/3</td>
<td>0/3</td>
<td>2/3</td>
<td>3/33</td>
<td>2/4</td>
<td>0/2</td>
<td>0/-</td>
</tr>
<tr>
<td>Koch J.A.</td>
<td>0/3</td>
<td>0/7</td>
<td>0/11</td>
<td>0/8</td>
<td>0/7</td>
<td>5/6</td>
<td>0/13</td>
<td>0/3</td>
<td>0/5</td>
<td>0/-</td>
</tr>
<tr>
<td>Kotze, T.N.</td>
<td>0/5</td>
<td>2/7</td>
<td>0/21</td>
<td>0/2</td>
<td>0/3</td>
<td>4/4</td>
<td>2/14</td>
<td>2/9</td>
<td>0/3</td>
<td>0/-</td>
</tr>
<tr>
<td>Krige, Leopold J.</td>
<td>7/7</td>
<td>-</td>
<td>3/3</td>
<td>7/7</td>
<td>4/4</td>
<td>11/11</td>
<td>8/8</td>
<td>1/1</td>
<td>1/1</td>
<td>0/-</td>
</tr>
<tr>
<td>Kuhn, Pieter G.</td>
<td>1/2</td>
<td>2/2</td>
<td>3/10</td>
<td>1/4</td>
<td>0/2</td>
<td>1/3</td>
<td>0/5</td>
<td>1/3</td>
<td>0/9</td>
<td>0/-</td>
</tr>
<tr>
<td>Laubscher, [Mimie]</td>
<td>1/1</td>
<td>13/13</td>
<td>22/22</td>
<td>7/7</td>
<td>8/8</td>
<td>18/18</td>
<td>31/32</td>
<td>5/5</td>
<td>1/10</td>
<td>11/682</td>
</tr>
<tr>
<td>Leipoldt, Carolina J.M.</td>
<td>15/45</td>
<td>22/49</td>
<td>3/121</td>
<td>7/44</td>
<td>1/43</td>
<td>90/104</td>
<td>3/235</td>
<td>2/17</td>
<td>0/73</td>
<td>3/597</td>
</tr>
<tr>
<td>LeRoux, D.J.</td>
<td>0/1</td>
<td>-</td>
<td>1/21</td>
<td>0/2</td>
<td>1/3</td>
<td>1/12</td>
<td>5/6</td>
<td>-</td>
<td>0/-</td>
<td>0/-</td>
</tr>
<tr>
<td>LeRoux, Petrus J. [1]</td>
<td>1/19</td>
<td>0/30</td>
<td>1/59</td>
<td>0/12</td>
<td>0/15</td>
<td>26/34</td>
<td>0/54</td>
<td>6/24</td>
<td>0/28</td>
<td>0/-</td>
</tr>
<tr>
<td>name</td>
<td>Gender</td>
<td>Verb(main)</td>
<td>Infinitive</td>
<td>PP - t</td>
<td>PP - en</td>
<td>Tense</td>
<td>Pronoun*</td>
<td>Om...te</td>
<td>Nie-2</td>
<td>Vir</td>
</tr>
<tr>
<td>-----------------------</td>
<td>--------</td>
<td>------------</td>
<td>------------</td>
<td>--------</td>
<td>---------</td>
<td>-------</td>
<td>----------</td>
<td>---------</td>
<td>-------</td>
<td>-----</td>
</tr>
<tr>
<td>Louw, Andries A. sen.</td>
<td>0/8</td>
<td>0/7</td>
<td>1/21</td>
<td>0/2</td>
<td>0/5</td>
<td>1/2</td>
<td>1/19</td>
<td>1/5</td>
<td>0/7</td>
<td>0/-</td>
</tr>
<tr>
<td>Louw, Andries A. jun.</td>
<td>0/11</td>
<td>0/12</td>
<td>0/16</td>
<td>0/7</td>
<td>0/7</td>
<td>9/13</td>
<td>1/38</td>
<td>3/8</td>
<td>0/6</td>
<td>0/-</td>
</tr>
<tr>
<td>Louw, Francina S.</td>
<td>0/4</td>
<td>0/13</td>
<td>0/21</td>
<td>0/5</td>
<td>0/6</td>
<td>3/4</td>
<td>0/35</td>
<td>2/6</td>
<td>0/9</td>
<td>1/591</td>
</tr>
<tr>
<td>Louw, Maria J.</td>
<td>0/2</td>
<td>0/4</td>
<td>1/18</td>
<td>0/4</td>
<td>0/1</td>
<td>2/3</td>
<td>0/12</td>
<td>1/9</td>
<td>0/6</td>
<td>0/-</td>
</tr>
<tr>
<td>Louw, Rijkie Hester</td>
<td>1/1</td>
<td>0/8</td>
<td>0/15</td>
<td>0/1</td>
<td>-</td>
<td>3/3</td>
<td>0/16</td>
<td>1/5</td>
<td>0/3</td>
<td>0/-</td>
</tr>
<tr>
<td>Louwrens, Anna S.</td>
<td>-</td>
<td>0/5</td>
<td>-</td>
<td>1/2</td>
<td>1/1</td>
<td>1/8</td>
<td>1/1</td>
<td>0/3</td>
<td>0/3</td>
<td>0/-</td>
</tr>
<tr>
<td>Lourens, Joh.F.</td>
<td>0/2</td>
<td>-</td>
<td>0/3</td>
<td>1/1</td>
<td>0/3</td>
<td>3/3</td>
<td>0/7</td>
<td>-</td>
<td>0/2</td>
<td>0/-</td>
</tr>
<tr>
<td>Malan, Daniel G.</td>
<td>0/5</td>
<td>0/23</td>
<td>0/29</td>
<td>0/15</td>
<td>0/3</td>
<td>10/15</td>
<td>1/34</td>
<td>0/8</td>
<td>0/11</td>
<td>0/-</td>
</tr>
<tr>
<td>Malan, Elizabeth J.</td>
<td>3/14</td>
<td>0/37</td>
<td>9/75</td>
<td>1/14</td>
<td>0/13</td>
<td>10/16</td>
<td>0/90</td>
<td>6/17</td>
<td>0/13</td>
<td>1/1753</td>
</tr>
<tr>
<td>Malan, Francois S. [1]</td>
<td>2/12</td>
<td>1/21</td>
<td>2/49</td>
<td>1/10</td>
<td>0/16</td>
<td>13/26</td>
<td>7/53</td>
<td>6/16</td>
<td>0/22</td>
<td>0/-</td>
</tr>
<tr>
<td>Malan, Francois S. [2]</td>
<td>42/42</td>
<td>41/44</td>
<td>96/111</td>
<td>18/19</td>
<td>14/16</td>
<td>31/31</td>
<td>98/117</td>
<td>23/30</td>
<td>36/50</td>
<td>17/2815</td>
</tr>
<tr>
<td>Malan, Jacques</td>
<td>7/7</td>
<td>3/4</td>
<td>2/2</td>
<td>1/2</td>
<td>2/3</td>
<td>5/5</td>
<td>20/24</td>
<td>0/1</td>
<td>0/4</td>
<td>1/328</td>
</tr>
<tr>
<td>Malan, Lelon</td>
<td>3/3</td>
<td>4/4</td>
<td>20/20</td>
<td>4/5</td>
<td>3/3</td>
<td>7/7</td>
<td>22/22</td>
<td>4/5</td>
<td>0/4</td>
<td>2/465</td>
</tr>
<tr>
<td>Malan, Maria C.</td>
<td>0/18</td>
<td>0/26</td>
<td>0/48</td>
<td>0/12</td>
<td>0/13</td>
<td>18/30</td>
<td>0/58</td>
<td>4/11</td>
<td>0/29</td>
<td>0/-</td>
</tr>
<tr>
<td>Malherbe, C.</td>
<td>1/1</td>
<td>1/1</td>
<td>2/6</td>
<td>-</td>
<td>-</td>
<td>0/1</td>
<td>1/4</td>
<td>4/5</td>
<td>0/1</td>
<td>0/-</td>
</tr>
<tr>
<td>Marais, F.P.</td>
<td>3/3</td>
<td>5/5</td>
<td>7/12</td>
<td>1/1</td>
<td>-</td>
<td>-</td>
<td>4/7</td>
<td>3/5</td>
<td>4/6</td>
<td>2/233</td>
</tr>
<tr>
<td>Marais, Jacobus P.Z.</td>
<td>0/10</td>
<td>0/8</td>
<td>0/25</td>
<td>0/10</td>
<td>0/15</td>
<td>13/20</td>
<td>0/31</td>
<td>3/11</td>
<td>0/9</td>
<td>0/-</td>
</tr>
<tr>
<td>Marais, Helena E.</td>
<td>6/8</td>
<td>4/15</td>
<td>5/26</td>
<td>0/1</td>
<td>1/7</td>
<td>8/12</td>
<td>7/34</td>
<td>6/6</td>
<td>0/4</td>
<td>1/564</td>
</tr>
<tr>
<td>Marais, Rachel</td>
<td>0/1</td>
<td>0/8</td>
<td>0/3</td>
<td>0/1</td>
<td>2/8</td>
<td>0/11</td>
<td>3/6</td>
<td>0/9</td>
<td>0/9</td>
<td>0/-</td>
</tr>
<tr>
<td>Marais, Stephanus J.</td>
<td>1/1</td>
<td>-</td>
<td>2/17</td>
<td>0/1</td>
<td>0/1</td>
<td>0/15</td>
<td>0/2</td>
<td>0/10</td>
<td>1/283</td>
<td>0/-</td>
</tr>
<tr>
<td>Marren, 'Maggie'</td>
<td>7/7</td>
<td>5/6</td>
<td>10/11</td>
<td>5/6</td>
<td>2/2</td>
<td>7/7</td>
<td>18/22</td>
<td>3/4</td>
<td>2/8</td>
<td>2/537</td>
</tr>
<tr>
<td>Matolana, Mozes</td>
<td>1/1</td>
<td>6/6</td>
<td>4/7</td>
<td>3/6</td>
<td>2/9</td>
<td>12/14</td>
<td>2/19</td>
<td>0/1</td>
<td>4/6</td>
<td>3/311</td>
</tr>
<tr>
<td>Mlimkulu, Mozes</td>
<td>-</td>
<td>0/5</td>
<td>3/26</td>
<td>1/7</td>
<td>0/4</td>
<td>6/7</td>
<td>0/32</td>
<td>5/9</td>
<td>0/5</td>
<td>3/575</td>
</tr>
<tr>
<td>Muller, Hermanus Steyn</td>
<td>1/1</td>
<td>3/4</td>
<td>21/23</td>
<td>1/2</td>
<td>1/3</td>
<td>4/5</td>
<td>2/11</td>
<td>9/9</td>
<td>0/9</td>
<td>2/431</td>
</tr>
<tr>
<td>November, Jacobus</td>
<td>1/1</td>
<td>-</td>
<td>3/5</td>
<td>0/1</td>
<td>-</td>
<td>1/1</td>
<td>0/2</td>
<td>2/2</td>
<td>0/1</td>
<td>1/100</td>
</tr>
<tr>
<td>Odendaal, Johanna</td>
<td>1/1</td>
<td>1/3</td>
<td>2/5</td>
<td>0/1</td>
<td>1/6</td>
<td>8/8</td>
<td>1/9</td>
<td>0/1</td>
<td>0/2</td>
<td>0/-</td>
</tr>
<tr>
<td>Oosthuizen, M. A.M.</td>
<td>3/17</td>
<td>5/19</td>
<td>5/34</td>
<td>3/7</td>
<td>0/3</td>
<td>7/13</td>
<td>15/89</td>
<td>6/13</td>
<td>4/16</td>
<td>1/1007</td>
</tr>
<tr>
<td>Pauw, Anna</td>
<td>0/2</td>
<td>0/8</td>
<td>0/13</td>
<td>0/2</td>
<td>-</td>
<td>3/8</td>
<td>1/14</td>
<td>2/3</td>
<td>0/5</td>
<td>0/-</td>
</tr>
<tr>
<td>Pienaar, Hendrik H.</td>
<td>1/1</td>
<td>1/1</td>
<td>12/18</td>
<td>2/2</td>
<td>1/4</td>
<td>4/4</td>
<td>1/16</td>
<td>2/4</td>
<td>2/3</td>
<td>1/347</td>
</tr>
</tbody>
</table>

329
<table>
<thead>
<tr>
<th>name</th>
<th>Gender</th>
<th>Verb(main)</th>
<th>Infinitive</th>
<th>PP - t</th>
<th>PP - en</th>
<th>Tense</th>
<th>Pronoun*</th>
<th>Om...te</th>
<th>Nie-2</th>
<th>Vir</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roos, ‘Lenie’</td>
<td>-</td>
<td>-</td>
<td>6/8</td>
<td>0/2</td>
<td>0/3</td>
<td>5/6</td>
<td>1/7</td>
<td>1/1</td>
<td>0/2</td>
<td>0/-</td>
</tr>
<tr>
<td>Roussouw, Aletta M.</td>
<td>1/2</td>
<td>2/2</td>
<td>2/5</td>
<td>0/2</td>
<td>0/2</td>
<td>5/5</td>
<td>0/18</td>
<td>0/1</td>
<td>-</td>
<td>0/-</td>
</tr>
<tr>
<td>Roux, Katie</td>
<td>3/4</td>
<td>4/9</td>
<td>6/14</td>
<td>1/5</td>
<td>0/2</td>
<td>7/7</td>
<td>15/21</td>
<td>3/8</td>
<td>0/4</td>
<td>0/-</td>
</tr>
<tr>
<td>Schabert, Elizabeth A.S.</td>
<td>48/51</td>
<td>45/72</td>
<td>56/139</td>
<td>12/27</td>
<td>8/51</td>
<td>27/69</td>
<td>29/283</td>
<td>8/26</td>
<td>0/75</td>
<td>2/4859</td>
</tr>
<tr>
<td>Schrade, Hester</td>
<td>3/3</td>
<td>1/1</td>
<td>7/7</td>
<td>3/3</td>
<td>3/3</td>
<td>5/5</td>
<td>10/10</td>
<td>4/4</td>
<td>4/5</td>
<td>0/-</td>
</tr>
<tr>
<td>Small, Nicolaas</td>
<td>3/3</td>
<td>9/11</td>
<td>11/16</td>
<td>2/2</td>
<td>7/12</td>
<td>12/12</td>
<td>12/41</td>
<td>3/3</td>
<td>0/7</td>
<td>5/534</td>
</tr>
<tr>
<td>Smit, ‘Bette’</td>
<td>36/37</td>
<td>61/66</td>
<td>57/76</td>
<td>30/32</td>
<td>36/41</td>
<td>83/84</td>
<td>190/204</td>
<td>17/22</td>
<td>25/50</td>
<td>10/3757</td>
</tr>
<tr>
<td>Smith, Edmund T.</td>
<td>5/7</td>
<td>7/9</td>
<td>17/25</td>
<td>7/8</td>
<td>3/8</td>
<td>20/20</td>
<td>10/36</td>
<td>5/8</td>
<td>0/10</td>
<td>1/756</td>
</tr>
<tr>
<td>Smuts, J.A. (‘Koos’)</td>
<td>0/2</td>
<td>0/1</td>
<td>2/10</td>
<td>0/1</td>
<td>-</td>
<td>2/8</td>
<td>5/17</td>
<td>1/4</td>
<td>0/2</td>
<td>0/-</td>
</tr>
<tr>
<td>Smuts, J.C.</td>
<td>0/6</td>
<td>0/2</td>
<td>1/13</td>
<td>0/2</td>
<td>0/4</td>
<td>6/13</td>
<td>0/11</td>
<td>1/7</td>
<td>0/5</td>
<td>0/-</td>
</tr>
<tr>
<td>Smuts, S. M. [1]</td>
<td>0/1</td>
<td>0/7</td>
<td>0/16</td>
<td>0/3</td>
<td>0/1</td>
<td>4/7</td>
<td>1/18</td>
<td>2/4</td>
<td>0/2</td>
<td>1/393</td>
</tr>
<tr>
<td>Smuts, S. M. [2]</td>
<td>3/3</td>
<td>3/3</td>
<td>5/5</td>
<td>1/1</td>
<td>2/3</td>
<td>3/3</td>
<td>2/6</td>
<td>1/1</td>
<td>-</td>
<td>0/-</td>
</tr>
<tr>
<td>Steyn, Anna</td>
<td>1/3</td>
<td>2/7</td>
<td>0/15</td>
<td>1/5</td>
<td>0/6</td>
<td>3/8</td>
<td>0/25</td>
<td>1/3</td>
<td>0/6</td>
<td>0/-</td>
</tr>
<tr>
<td>Steyn, Daniel</td>
<td>1/1</td>
<td>2/2</td>
<td>4/6</td>
<td>3/3</td>
<td>1/3</td>
<td>4/4</td>
<td>3/9</td>
<td>3/3</td>
<td>-</td>
<td>0/-</td>
</tr>
<tr>
<td>Steyn, Elizabeth G.</td>
<td>10/12</td>
<td>9/11</td>
<td>14/28</td>
<td>8/10</td>
<td>5/8</td>
<td>18/18</td>
<td>14/53</td>
<td>6/8</td>
<td>0/9</td>
<td>7/1100</td>
</tr>
<tr>
<td>Steyn, Rachel I.</td>
<td>15/15</td>
<td>29/29</td>
<td>57/59</td>
<td>25/27</td>
<td>9/9</td>
<td>37/38</td>
<td>73/75</td>
<td>11/13</td>
<td>7/22</td>
<td>5/1613</td>
</tr>
<tr>
<td>Uys, Susan</td>
<td>2/2</td>
<td>1/1</td>
<td>4/9</td>
<td>2/4</td>
<td>1/2</td>
<td>6/7</td>
<td>2/9</td>
<td>-</td>
<td>0/4</td>
<td>0/-</td>
</tr>
<tr>
<td>vd Merwe, C.J.</td>
<td>3/3</td>
<td>5/6</td>
<td>5/30</td>
<td>0/2</td>
<td>0/4</td>
<td>5/5</td>
<td>10/21</td>
<td>3/12</td>
<td>0/13</td>
<td>1/667</td>
</tr>
<tr>
<td>vHuyssteen, A.J.Z. (‘Janie’)</td>
<td>1/2</td>
<td>0/4</td>
<td>2/15</td>
<td>1/4</td>
<td>0/2</td>
<td>7/9</td>
<td>1/13</td>
<td>1/2</td>
<td>0/2</td>
<td>0/-</td>
</tr>
<tr>
<td>vHuyssteen, Anna Chr.</td>
<td>1/2</td>
<td>0/1</td>
<td>2/15</td>
<td>-</td>
<td>0/1</td>
<td>0/10</td>
<td>1/13</td>
<td>0/6</td>
<td>0/2</td>
<td>0/-</td>
</tr>
<tr>
<td>vHuyssteen, Dan. P.S.</td>
<td>1/4</td>
<td>0/4</td>
<td>3/24</td>
<td>0/2</td>
<td>0/2</td>
<td>3/4</td>
<td>0/23</td>
<td>0/4</td>
<td>0/4</td>
<td>0/-</td>
</tr>
<tr>
<td>vHuyssteen, Eliza J.</td>
<td>5/11</td>
<td>1/12</td>
<td>0/48</td>
<td>0/13</td>
<td>0/11</td>
<td>20/45</td>
<td>0/64</td>
<td>9/15</td>
<td>0/30</td>
<td>1/1356</td>
</tr>
<tr>
<td>vHuyssteen, E.C.P.</td>
<td>17/29</td>
<td>40/49</td>
<td>34/100</td>
<td>22/32</td>
<td>0/32</td>
<td>67/72</td>
<td>40/185</td>
<td>17/34</td>
<td>0/60</td>
<td>4/4304</td>
</tr>
<tr>
<td>vHuyssteen, Hendrik</td>
<td>3/6</td>
<td>10/13</td>
<td>14/41</td>
<td>5/5</td>
<td>0/4</td>
<td>5/5</td>
<td>7/28</td>
<td>9/9</td>
<td>0/7</td>
<td>3/787</td>
</tr>
<tr>
<td>vHuyssteen, J.P.J.W.</td>
<td>1/16</td>
<td>0/12</td>
<td>0/33</td>
<td>0/10</td>
<td>0/13</td>
<td>15/45</td>
<td>0/62</td>
<td>1/11</td>
<td>0/15</td>
<td>0/-</td>
</tr>
<tr>
<td>vHuyssteen, Jacomina H.</td>
<td>4/4</td>
<td>5/17</td>
<td>3/34</td>
<td>0/8</td>
<td>0/3</td>
<td>8/78</td>
<td>7/59</td>
<td>0/11</td>
<td>0/16</td>
<td>0/-</td>
</tr>
<tr>
<td>vHuyssteen, M.C. (‘Katie’)</td>
<td>1/14</td>
<td>0/18</td>
<td>2/68</td>
<td>0/6</td>
<td>0/9</td>
<td>15/17</td>
<td>1/55</td>
<td>16/19</td>
<td>0/21</td>
<td>0/-</td>
</tr>
<tr>
<td>vHuyssteen, Petrus H.H.</td>
<td>0/6</td>
<td>0/8</td>
<td>0/22</td>
<td>0/5</td>
<td>0/4</td>
<td>10/17</td>
<td>6/44</td>
<td>10/14</td>
<td>0/12</td>
<td>0/-</td>
</tr>
<tr>
<td>vHuyssteen, Matthys P.Z.</td>
<td>0/1</td>
<td>1/5</td>
<td>0/9</td>
<td>0/2</td>
<td>0/2</td>
<td>3/3</td>
<td>0/11</td>
<td>0/6</td>
<td>0/1</td>
<td>0/-</td>
</tr>
</tbody>
</table>
Appendix B 4: Cluster Membership

Figure 8.3.

Cluster I

M. Basson
Catharina E. Beyers [1]
Anna De Vries
Jan H.H. De Waal [1]
P.A. Goosen, jun. [1]
A. Hauptfleich [1]
F.P. Hoogenhout [1]
N.M. Hoogenhout
Wynanda Hoogenhout [1]
Carolina Leipoldt
P.J. LeRoux [1]
A.A. Louw, jun.
Elizabeth J. Malan
F.S. Malan [1]
Maria Malan
J.P.Z. Marais
J.P.J.W. Van Huyssteen
Eliza J. Van Huyssteen
M.C. ('Katie') Van Huyssteen

Cluster II

Louis Botha
B.J. Brümmer
Elizabeth Eksteen
A.S. Schabert
Elizabeth Steyn
E.C.P. Van Huyssteen
H.S. Vermaak
E. Smith

Cluster III

Johanna Brümmer
D.C. De Wet
G.B. Goosen
Sebastina Goosen
J.F.W. Grosskopf
Hester Hoogenhout
Wynanda Hoogenhout [2]
F.S. Malan [2]
'Bettie' Smit
Rachel I. Steyn

Figure 8.9

Cluster I

M. Basson
Catharina E. Beyers [1]
Anna De Vries
Jan H.H. De Waal [1]
P.A. Goosen, jun. [1]
A. Hauptfleich [1]
F.P. Hoogenhout [1]
N.M. Hoogenhout
Wynanda Hoogenhout [1]
Carolina Leipoldt
P.J. LeRoux [1]
A.A. Louw, jun.
Elizabeth J. Malan
F.S. Malan [1]
Maria Malan
J.P.Z. Marais
J.P.J.W. Van Huyssteen
Eliza J. Van Huyssteen
M.C. ('Katie') Van Huyssteen

Cluster II

W.C. Cronje
B.J. Brümmer
F.S. Cillie
A.S. Schabert
Elizabeth Steyn
E.C.P. Van Huyssteen
B. Hattingh
M. Matodlana
Johanna Brümmer [2]
Engela J.M. Furter
E. Smith

Cluster III

C. Coopman
D.C. De Wet
G.B. Goosen
Sebastina Goosen
J.F.W. Grosskopf
Hester Hoogenhout
Wynanda Hoogenhout [2]
F.S. Malan [2]
'Bettie' Smit
Rachel I. Steyn
E. Eksteen
M. Laubscher
D.M. Hoogenhout [2]
Cornelia Stanford
H.S. Vermaak
BIBLIOGRAPHY

MANUSCRIPT SOURCES
CA  THE STATE ARCHIVES DEPOT, CAPE TOWN
    A 583  F.S. Malan Collection

SAL  SOUTH AFRICAN LIBRARY MANUSCRIPT COLLECTION
    MSC 1  Hoogenhout Collection
    MSC 7  J.H. De Villiers
    MSC 33  Rae Steyn Collection
    MSB 162  Du Preez Family Collection
    MSB 229  Heroldt Family Papers
    MSB 437  A.S. Schabort
    MSB 510  Volkscongres, Worcester (1900)
    MSB 744  South Africa Letters (Miscellaneous Collection)
    MSB 852  D.J. Malan Collection
    MSB 853  Katie Malan Collection
    MSB 892  Women's Studies (Miscellaneous Collection)
    MSB 957  Mimie Reyneke Collection

UCT  MANUSCRIPT COLLECTION OF THE UNIVERSITY OF CAPE TOWN
    B28  De Villiers Collection
    BCZA 77  Smuts Collection
    BC94  Louis Leipoldt Papers

NGK  ARCHIVES OF THE NEDERLANDS GEREFORMEERDE KERK, CAPE TOWN
    P  Private Papers
    S 5  Sendingskerk

G  GENADENDAL ARCHIVES

OFFICIAL PUBLICATIONS
CPP  CAPE PARLIAMENTARY PAPERS
    G42-1876  Cape Census for 1875
    G6 - 1892  Cape Census for 1891
    G19- 1905  Cape Census for 1904

G75-1880  Report of the Commission appointed to inquire into and report upon the working of the Education Acts in force in this Colony. 1879-1880.

G12-1883  Preliminary Report on the State of Education in the Colony of the Cape of Good Hope. (Ross-Report)

G3 1892  Education Commission, First and Second Report.

G 682 E1-1892  Third and final Report and Proceedings, with Appendices, of a Commission appointed to inquire into and report upon certain matters connected with the educational system of the Colony.

UNION OF SOUTH AFRICA

UG32-1912  Census of the Union of South Africa, 1911
VOTER'S LIST AND DIRECTORIES

PERIODICALS
Die Afrikaanse Patriot (Paarl)
Die Burger (Cape Town)
The Cape Monthly Magazine (Cape Town).
The Cape Times (Cape Town)
Ons Land (Cape Town)
De Zuid Afrikaan (Cape Town)
Het Zuid-Afrikaansche Tijdschrift (Cape Town)
De Goede Hoop (Cape Town)
De Ebenezer (Cape Town)
De Bode van Genadendal (Genadendal)

BOOKS AND JOURNALS (SECONDARY SOURCES)
An. (1873): Benigna van Groenekloof of Mambre. Een Verhaal voor de Christen Kleuringen van Zuid-
Amsterdam: John Benjamins, 1-20.