READABILITY AND GERMAN BIBLES

A Dissertation in

German

by

Nathan L. Shrefler

© 2008 Nathan L. Shrefler

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2008
The Dissertation of Nathan L. Shrefler was reviewed and approved* by the following:

Barry Richard Page, Jr.
Associate Professor of German and Linguistics
Dissertation Advisor
Chair of Committee

Carrie Neal Jackson
Assistant Professor of German and Linguistics

Steven Leadley Thorne
Assistant Professor of Applied Linguistics

Sinfile Bullock Makoni
Associate Professor of Applied Linguistics and African Studies

Daniel Leonhard Purdy
Associate Professor of German
Chair of Graduate Program of the Department of Germanic and Slavic Languages and Literatures

*Signatures are on file in the Graduate School
The primary goal of this dissertation is to compare two German Bible versions in terms of readability and related linguistic features: the classic translation of Martin Luther and the more modern translation Hoffnung für alle. Both versions, at the times of their publishing, claimed to be done in a style consistent with the ways in which contemporary Germans used their language – that they were, in effect, very readable. I will be focusing specifically on aspects of the texts that contribute or detract from the ease of reading.

I examine identical sections from the two Bible versions, including the four New Testament Gospels and the Acts of the Apostles. In order to compare the readability of these two texts, I explore four linguistic areas via the use of corpus analysis: 1) lexical density, 2) clausal complexity, 3) nominalization, and 4) lexical bundles. Each of these areas is analyzed based on markers that can be searched for within the corpus, relying mainly on part-of-speech tagging of content words and conjunctions, suffix markers of nominalization, and the frequency of recurring lexical bundles. Using the patterns established via corpus analysis, I discuss the implications of those patterns for readability.

There has been much criticism and praise of both versions of the Bible to be discussed here. Most criticism, however, is based on limited linguistic data. It is the aim of this project to compare the two versions via statistically prevalent linguistic aspects of readability found in the two texts. Critique based on a systematic study of readability will contribute to the vast literature already present on Bible translation, as well as to the literature engaged in the study of Luther and his works. Results show that Luther was the more readable text for its contemporaries from all aspects of statistical analysis used here.
TABLE OF CONTENTS

LIST OF FIGURES ..................................................................................................... vii
LIST OF TABLES ....................................................................................................... ix
ACKNOWLEDGEMENTS ......................................................................................... xi
Chapter 1  Introduction .......................................................................................... 1
  1.1 Readability and German Bibles ................................................................. 1
  1.2 Research Questions ....................................................................................... 2
  1.3 The Importance of Luther ................................................................. 3
     1.3.1 Luther and the German language ......................................................... 5
  1.4 The Importance of Hoffnung für alle .......................................................... 8
  1.5 Literacy, Orality and Religiosity ............................................................ 12
  1.6 Readability ................................................................................................ 14
     1.6.1 Who Studies Readability ...................................................................... 15
  1.7 Motivation for Study .................................................................................... 17
Chapter 2  Background and Methodology .......................................................... 19
  2.1 Introduction ................................................................................................. 19
  2.2 Corpus Analysis .......................................................................................... 20
  2.3 Readability ................................................................................................ 24
     2.3.1 Readability Formulas ........................................................................... 25
     2.3.2 Criticism of Readability Formulas ....................................................... 30
  2.4 Lexical Density ........................................................................................... 31
  2.5 Clausal Complexity ....................................................................................... 33
  2.6 Nominalization ........................................................................................... 35
  2.7 Lexical Bundles .......................................................................................... 41
  2.8 Conclusion .................................................................................................. 45
Chapter 3  Lexical Density .................................................................................... 46
  3.1 Introduction ................................................................................................. 46
  3.2 Methodology ............................................................................................... 46
     3.2.1 Corpus ................................................................................................ 47
     3.2.2 Part of Speech Tagging .......................................................................... 47
     3.2.3 Lexical Density Calculation ................................................................. 48
     3.2.4 Noun and Verb Concentration ............................................................. 49
  3.3 Data Analysis .............................................................................................. 50
     3.3.1 Lexical Density .................................................................................... 50
     3.3.2 Lexical Word Types ............................................................................ 60
     3.3.2.1 Concentration of Nouns ............................................................ 60
### Table of Contents

3.3.2.2 Concentration of Verbs .............................................................. 68  
3.4 Discussion....................................................................................................... 76  
3.4.1 Lexical Density Patterns....................................................................... 76  
3.4.2 Differences in Types of Lexical Words................................................... 78  
3.5 Conclusion...................................................................................................... 80  

Chapter 4 Clausal Complexity.................................................................................... 81  
4.1 Introduction..................................................................................................... 81  
4.2 Methodology................................................................................................... 82  
4.3 Data Analysis.................................................................................................. 83  
4.3.1 Frequency of Different Conjunction Types.......................................... 83  
4.3.2 Comparative Frequency of Coordinating Conjunctions....................... 89  
4.3.3 Comparative Frequencies of Subordinating Conjunctions................... 94  
4.3.4 Correlation Analysis of Conjunction Frequency and Lexical Density ................................................................................................... 97  
4.4 Discussion....................................................................................................... 98  
4.5 Conclusion ...................................................................................................... 100  

Chapter 5 Nominalization........................................................................................... 102  
5.1 Introduction..................................................................................................... 102  
5.2 Methodology................................................................................................... 103  
5.2.1 Statistical Analysis ............................................................................... 103  
5.2.2 Analysis of Nominalizations in Context............................................... 105  
5.3 Data Analysis.................................................................................................. 107  
5.3.1 Statistical Analysis of Nominalization Types ...................................... 107  
5.3.2 Analysis of Nominalizations in Context............................................... 111  
5.3.2.1 Auferstehung ............................................................................... 111  
5.3.2.2 Gotteslästerung .......................................................................... 114  
5.3.2.3 Versuchung ................................................................................. 114  
5.3.2.4 Verheißung ............................................................................... 116  
5.3.2.5 Anfectung .................................................................................... 117  
5.4 Discussion....................................................................................................... 118  
5.4.1 Comparison of the Most Frequent Nominalizations in Luther and Hfa.......................................................................................................... 118  
5.4.2 Comparison of Nominalizations in Context .......................................... 121  
5.4.2.1 Overlapping Nominalizations .................................................... 121  
5.4.2.2 Non-overlapping Nominalizations .............................................. 122  
5.5 Conclusion ...................................................................................................... 125  

Chapter 6 Lexical Bundles.......................................................................................... 126  
6.1 Introduction..................................................................................................... 126  
6.2 Methodology................................................................................................... 127  
6.2.2 Functions of Lexical Bundles............................................................... 128
6.3 Data Analysis ........................................................................................................ 129
  6.3.1 Statistical Comparison of Lexical Bundle Use ........................................ 129
  6.3.2 Lexical Bundles in Context ....................................................................... 132
    6.3.2.1 Type Trends in Four-word Bundles ............................................. 132
    6.3.2.2 Comparison of Functions of Most Frequent Four-word Bundles ....... 134
      6.3.2.2.1 Luther: und sprach zu ihnen ............................................. 135
      6.3.2.2.2 Luther: der war ein Sohn ............................................. 136
      6.3.2.2.3 Luther: wahrlich ich sage euch ........................................ 138
      6.3.2.2.4 Luther: und er sprach zu .................................................. 139
      6.3.2.2.5 Luther: antwortete und sprach zu ....................................... 141
      6.3.2.2.6 Hfa: der mich gesandt hat ................................................ 142
      6.3.2.2.7 Hfa: vor ihm nieder und .................................................. 143
      6.3.2.2.8 Hfa: Jesus mit seinen Jüngern ......................................... 144
      6.3.2.2.9 Hfa: in der heiligen Schrift ............................................... 146
      6.3.2.2.10 Hfa: was ich euch sage .................................................... 147
  6.4 Discussion .......................................................................................................... 148
    6.4.1 Comparative Frequency and Variety of Lexical Bundles ....................... 148
    6.4.2 Comparative Types and Functions of Lexical Bundles ........................ 149
    6.4.3 Lexical Bundles and Readability .......................................................... 154
  6.5 Conclusion ......................................................................................................... 155

Chapter 7 Conclusion .............................................................................................. 157
  7.1 Introduction ..................................................................................................... 157
  7.2 Summary of Findings ..................................................................................... 157
    7.2.1 Lexical Density ..................................................................................... 157
    7.2.2 Clausal Complexity .............................................................................. 159
    7.2.3 Nominalization ..................................................................................... 160
    7.2.4 Lexical Bundles .................................................................................... 161
  7.3 Implications of Findings for Readability ....................................................... 162
  7.4 Avenues for Further study .............................................................................. 165

Bibliography ............................................................................................................ 169

Appendix A Stuttgart-Tübingen Tagset (Schiller, Teufel and Stöckert, 1995) ........ 175
Appendix B Lexical Density Calculation Program ................................................. 176
  B.1 Lexical Density Calculator with all Verbs................................................... 177
  B.2 Lexical Density Calculator with Modals and Auxiliaries Excluded .......... 179

Appendix C The grep function in UNIX and its use in this project ................. 181
LIST OF FIGURES

Figure 2.1: Flesch Reading Ease Chart  (Flesch 1949, inside cover) .........................27
Figure 2.2: Fry Graph (http://school.discovery.com/schrockguide/fry/fry2.html accessed 8-27-2007) .............................................................................................28
Figure 3.1: Lexical Density per Verse ........................................................................52
Figure 3.2: Lexical Density per Verse without Modals and Auxiliaries ...............53
Figure 3.3: Lexical Density per Chapter ....................................................................55
Figure 3.4: Lexical Density per Chapter without Modals and Auxiliaries ..........56
Figure 3.5: Lexical Density per book and overall ..............................................57
Figure 3.6: Lexical Density per Book and Complete Corpus without Modals and Auxiliaries .................................................................59
Figure 3.7: Percent Nouns per Verse ......................................................................61
Figure 3.8: Percent Nouns per Verse without Modals and Auxiliaries ...............63
Figure 3.9: Percent Nouns per Chapter .................................................................64
Figure 3.10: Percent Nouns per Chapter without Modals and Auxiliaries ..........65
Figure 3.11: Percent Nouns per Book and Entire Corpus ....................................66
Figure 3.12: Percent Nouns per Book and Complete Corpus without Modals and Auxiliaries .................................................................68
Figure 3.13: Percent Verbs per Verse .......................................................................69
Figure 3.14: Percent Verbs per Verse without Modals and Auxiliaries ...............71
Figure 3.15: Percent Verbs per Chapter .................................................................72
Figure 3.16: Percent Verbs per Chapter without Modals and Auxiliaries ..........73
Figure 3.17: Percent Verbs per Book and Entire Corpus ....................................74
Figure 3.18: Percent Verbs per Book and Entire Corpus without Modals and Auxiliaries .................................................................75
Figure 4.1: Coordinating Conjunctions per Chapter ..............................................90
Figure 4.2: Coordinating Conjunctions per Chapter without und.................................91
Figure 4.3: Coordinating Conjunctions per Book..............................................................93
Figure 4.4: Coordinating Conjunctions per Book without und.........................................94
Figure 4.5: Subordinating Conjunctions per Chapter ......................................................95
Figure 4.6: Subordinating Conjunctions per Book ..........................................................96
Figure 5.1: Use of Nominalization in Secondary School History Texts (Unsworth 1999)...............................................................................................................................105
Figure B.1: Lexical Density Stat Program Output..............................................................176
LIST OF TABLES

Table 2.1: Parataxis (Halliday and Matthiessen, 2004, p. 387) ..................................34
Table 2.2: Hypotaxis (Halliday and Matthiessen, 2004, p. 387) ............................34
Table 2.3: Nominalstil vs. Verbaler Stil (Stedje, 1989, p. 181) .............................39
Table 2.4: Comparison of Top Seven Suffixes According to Type and Token (Russ, 2004, p. 252) .............................................................40

Table 3.1: Lexical Density Per Verse ......................................................................51
Table 3.2: Lexical Density per Verse without Modals and Auxiliaries .....................53
Table 3.3: Lexical Density per Chapter ..................................................................54
Table 3.4: Lexical Density per Chapter without Modals and Auxiliaries ...............56
Table 3.5: Lexical Density Comparison per book and the complete text .................57
Table 3.6: Lexical Density per Book and Complete Corpus without Modals and Auxiliaries ..........................................................58
Table 3.7: Percent Nouns per Verse ......................................................................61
Table 3.8: Percent Nouns per Verse without Modals and Auxiliaries .....................62
Table 3.9: Percent Nouns per Chapter ..................................................................63
Table 3.10: Percent Nouns per Chapter without Modals and Auxiliaries .......... 64
Table 3.11: Percent Nouns per Book and Entire Corpus .........................................66
Table 3.12: Percent Nouns per Book and Complete Corpus without Modals and Auxiliaries ..........................................................67
Table 3.13: Percent Verbs per Verse ....................................................................69
Table 3.14: Percent Verbs per Verse without Modals and Auxiliaries ....................70
Table 3.15: Percent Verbs per Chapter ..................................................................71
Table 3.16: Percent Verbs per Chapter without Modals and Auxiliaries ...............73
Table 3.17: Percent Verbs per Book and Entire Corpus .........................................74
Table 3.18: Percent Verbs per Book and Entire Corpus without Modals and Auxiliaries

Table 4.1: Coordinating Conjunctions in Luther

Table 4.2: Coordinating Conjunctions in Hfa

Table 4.3: Subordinating Conjunctions in Luther

Table 4.4: Subordinating Conjunctions in Hfa

Table 4.5: Coordinating Conjunctions per Chapter

Table 4.6: Coordinating Conjunctions per Chapter without und

Table 4.7: Coordinating Conjunctions per Book

Table 4.8: Coordinating Conjunctions per Book without und

Table 4.9: Subordinating Conjunctions per Chapter

Table 4.10: Subordinating Conjunctions per Book

Table 4.11: Overall Conjunction Frequency

Table 4.12: Correlation Coefficients for all Categories

Table 5.1: Ten most frequent –ung nominalizations

Table 5.2: Ten most frequent –heit nominalizations

Table 5.3: Ten most frequent –keit nominalizations

Table 5.4: Ten most frequent –nis nominalizations

Table 5.5: Nominalization Comparison

Table 6.1: Statistical Comparison of Lexical Bundles

Table 6.2: Expanding Bundles in Luther

Table 6.3: Expanding Bundles in Hfa

Table 6.4: Twenty most frequent four-word bundles in Luther

Table 6.5: Twenty most frequent four-word bundles in Hfa
ACKNOWLEDGEMENTS

I wish to thank the Institute for Arts and Humanities for their financial and creative support of this dissertation. The dissertation grant and semester teaching release were invaluable to the progress of my research. Primary support has come from the Department of Germanic and Slavic Languages and Literatures, and I thank all the staff of the department for their assistance during my years here at Penn State.

Many thanks go to my dissertation advisor Richard Page for his support and guidance over my eleven years of association with Penn State. I would also like to thank my committee members Carrie Jackson, Sinfree Makoni and Steven Thorne for their helpful input, support and excitement for my project.

My family has my thanks for their love, to my parents Larry and Marcheta for pushing me along this path and David and Carol Moore for welcoming me into their home and family.

Todd Ellis has my eternal gratitude for help, both practical and emotional, through not one, but two theses of mine at Penn State. His friendship, kindness and understanding have seen me through many things.

Most of all, my deepest thanks and love go to my wife, Amanda Moore. When many had seen this project as little more than a pipe dream, she saw and fostered the desire in me to complete this work and follow my dreams. This dissertation is a testament to her belief in me, and to the strength I found in myself through that belief.
Chapter 1

Introduction

1.1 Readability and German Bibles

Bible translation has been at the forefront of linguistic inquiry and controversy since ancient times. From Roman times through the Renaissance to the present there have been debates ranging from the translational equivalents of individual words through to entire phrases and structures. In modern times, the Bible has become one of the most highly translated and universally accessible books in the world, and the act of translation has become far more accepted, in comparison to the time of the Protestant Reformation, for example. Where the choice of Bible was once a highly charged cultural and political choice, today it is more a matter of consumer preference, although still not without cultural, political or social import. Dewey (2004) states that “[t]hose purchasing an untried Bible have two main concerns: is it reliable, and is it readable?” (p. 33) While the reliability of any given Bible is beyond the purview of this project, it is the second aspect of the above question – readability – that is at the heart of this project. The primary goal of this dissertation is to compare two German Bible versions in terms of readability and related linguistic features: the classic translation of Martin Luther (hereafter referred to as Luther) and the more modern translation Hoffnung für alle (Hope for All – hereafter Hfa). This project contributes to research on Martin Luther and his contributions to the German language and Christianity, as well as to research on modern translations of the Bible and
how they are received by their readers. This research applies methodologies used primarily with academic texts to the question of readability and the Bible, providing a new perspective to the question of what makes an important text like this accessible or not to its readers. The goal of this project is to apply statistical measures of readability to the assertion that both Luther and Hfa were produced with the intent of providing their contemporaries with a readable translation of the Bible. I utilize statistical aspects of corpus analysis in order to ascertain which version was more successful in this attempt.

1.2 Research Questions

1. Do Luther and Hfa differ in readability as assessed by a range of linguistic measures? If so, how do they differ?

This portion of the project will focus on specific linguistic features of both texts which figure prominently in the measure of readability. The specifics of this will be discussed later in this work, but it is important here to note the answering of this question will involve the cataloguing and comparison of a network of different items, from those traditionally associated with readability to those more indirectly linked to the concept. As I will show in the discussion of the two texts below, both versions were created with the goal of producing a Bible that was readable and accessible to the contemporaries of the translators.

2. Are the linguistic differences identified by measures of readability driven by linguistic changes in the German lexicon and syntax since the mid-16th century or by changes in the readership and use of the Bible since that time?
This question will deal with the synthesis of the results of the statistical measures used to answer question 1. Question 2 focuses more on the nature of the two texts and their reception by contemporary readers, asking whether the most important change is in the German language itself, or in the German-speaking people who read the two texts. Three areas will be of importance for the synthesis of the quantitative data from question 1: literacy, orality, and religiosity. Each of these areas will be discussed later in this chapter.

1.3 The Importance of Luther

There are few figures as important in German history as Martin Luther. His influence spans religion, language and culture, altering how people viewed God, the Church, race and class distinctions, and the German language itself. Luther was not the first to translate the Bible into German. By the time his Bible had been published there were already 14 printed versions in High German and 4 in Low German, as well as other translated sources (von Polenz, 2000, p. 230). The popularity of Luther’s translation, however, is quite amazing for the time period: the last version of the Bible that Luther participated in editing was completed in 1546 – by 1574 there were 100,000 copies in print, not including reprints (p. 134). Young and Gloning (2004) put the importance of Luther into perspective in this way:

Martin Luther is significant in German political and cultural history in two respects. He was the founder of the Reformation and, as a figure who communicated his opinions to the masses, had far-reaching – if still hotly debated – effects on the development of the German language from the early sixteenth century onwards. From the viewpoint of language history,
both factors – historical/biographical and linguistic – must be seen as interlocking. (p. 205).

It was this ability of Luther’s to reach such masses and have such a widespread effect on the German language that make his work important, even today. As can be seen in the quote above, Luther’s success occurred within a nexus of many different factors in linguistic, cultural and religious landscape of his time. His ability to use such factors to his advantage has allowed his work to stand the test of time. This fact is attested in the following quote from von Polenz (2000):

*Luthers kirchensprachliche Leistung war kein Neuanfang, sondern der Kulminationspunkt langfristiger Sprachtraditionen, in denen er aufgewachsen, die er sich angeeignet, aus denen er mit großer Sprachmeisterschaft geschöpft hat. Schon aus der Tatsache, daß er erst im Alter von 34 Jahren vom lateinischen zum (sogleich sehr gekonnten, erfolgreichen) deutschen Schreiben und Publizieren überging, ist zu schließ, daß er auf ihm längst geläufige deutschsprachige Vorbilder zurückgreifen konnte. In diesen Traditionen waren die bis heute gerühmten Eigenschaften des ‚Lutherdeutsch’ bereits weitgehend vorhanden: einfach, den Laien verständlich, gut sprechbar und hörbar, bildhaft, drastisch (p. 229).*

Luther’s church-language effort was no new beginning, but rather the culmination point of long-running language traditions, in which he grew up, that he adopted, and out of which he created with great language mastery. It can already be assumed from the fact that at age 34 he switched from Latin to (at the same time very refined and successful) German writing and publishing, that he could reach back to already long-running German-speaking examples. In these traditions the still-famous “Luther German” aspects were already widely present: simplicity, understandability to the laity, easily spoken and heard, full of imagery, drastic. (translation mine)

The Evangelische Kirche in Deutschland (EKD), comprised of “mainline” Protestant churches in the Lutheran and Reformed traditions, continues to use Luther’s translation today as its primary text (Lutherbibel, p. 4), with the newest revision to the text having taken place in 1984 (p. 5). Changes from the original and earlier revisions are
primarily orthographic in nature, with some words being replaced due to differing connotations between the present and Luther’s time (ibid.). That so little should have changed over the centuries is a testament to the quality and vividness of Luther’s work.

### 1.3.1 Luther and the German language

Von Polenz (2000), points out that Luther’s language work had the interesting characteristic of turning towards spoken German (p. 230). Many forms of spoken language appeared in his writing – including free word order, ellipsis, modal particles, and idioms (ibid.). Waterman (1966) presents a similar viewpoint on Luther’s translation:

*The Lutherbibel still stands as the most magnificent literary monument in the German language. Though its orthography and grammar are based to a large extent upon the chancery practices in the Saxon electorate, its style certainly is not. Luther’s great genius lay in his uncanny and inspired ability to forge a language that was idiomatic and natural (Waterman, 1966, p. 130).*

According to von Polenz (2000), Luther’s striving towards a spoken style of translation has to do with Luther’s identification with the language of preaching, rather than of academic or theological writing (p. 231). Luther himself states his view of the nature of the Gospel in this way:

*Das Evangelium ist eygentlich nicht das/ das ynn bechern stehet und ynn buchstaben verfasset wirtt/ sondern mehr eyn muendliche predig und lebendig wortt und eyn stym / die da ynn die gantz wellt erschallet und öffentlich wirt außgeschryen / das mans ueberal hoeret (Luther, quoted in von Polenz, 2000, p. 231).*

The Gospel is actually not that/ which is contained in books and composed in letters/ rather more an oral sermon and living word and a voice/ that
there in the entire world is sounding and is publicly cried out/ so that one
hears it everywhere. (translation mine)

Clearly, Luther viewed the language of the Bible not as that of a book, but as a
living voice to be heard aloud, and his translation of the Bible reflects this. Luther’s
goals in translating the Bible can be clearly seen in the following quote from his
Sendbrief vom Dolmetschen (‘Letter on Translation’, 1530, quoted in Young and
Gloning, 2004, p. 211):

Ich hab mich des geflissen ym dolmetschen/ das ich rein und klar teusch
ggeben möchte ... den man muß nicht die buchstaben inn der lateinischen
sprachen fragen.

Man mus die mutter jhm hause/ die kinder auff der gassen/ den gemeinen
man auff dem marckt drumb fragen/ vnd den selbigen auff das maul sehen/
wie sie reden/ vnd darnach dolmetszschen/ so verstehen sie es den/ vnd
mercken/ das man Deutsch mit jn redet.

I have disciplined myself in translation, that I would like to give clean and
clear German... because one need not ask the letters in the Latin language.

One must ask around to the mother in the house, the children on the street,
the regular man in the market, and look those same people in the face and
see how they speak and translate following that so that they understand
and realize that one is speaking German with them. (translation mine)

Luther sought, according to von Polenz (2000), to translate the Bible in a manner that
was based on the sense or message of the text, rather than a word-for-word translation of
the original Greek or the Latin Vulgate (p. 231).

In this endeavor, Luther was not without detractors. For example, Hieronymus
Emser attacked Luther’s translation on three levels: 1) that Luther’s translation was not
authorized by the Church; 2) that Luther ignored the Latin Vulgate in his translation; and 3) that Luther translated things incorrectly (Gelhaus, 1989).

Emser sought to uphold the authority of the Catholic Church over the language of the Bible. Ironically, however, Emser himself produced a “translation” which earned him his nickname as the Scribbler of Dresden. Furthermore, when laid side by side with Luther’s translations, the two are almost identical (p. 53).

In a similar argument, Witzel condemns Luther’s use of poetic license when translating the Bible (Gelhaus, 1989, p. 68). While this may have been good enough for translating Homer or other ancient texts, such translation was not, in Witzel’s mind, appropriate in the case of the Bible (p. 69). Specifically, Witzel criticized Luther for adding, removing, and varying words from previous versions. Witzel also pointed out places where Luther made changes to tense, person, and case (p. 75).

Despite such criticisms, Luther strove to create a Bible translation that was clearly meant to be dynamic, language that was meant to be heard and not simply read, language that was meant to be living and vital, cried out, as it were, in all places. It is this nature of the text that makes it a fascinating choice as one of the Bibles to be compared in this study. Clearly, from all the statements above, Luther approached his work with a clear goal in mind – to produce a Bible that was understandable to all Germans of the time – and he strove toward this goal for much of his life, and which is continued today by those taking part in the continued revisions of his text.
1.4 The Importance of Hoffnung für alle

The choice of a modern Bible for comparison to Luther is more difficult to justify than the choice of Luther’s translation. Almost any translation from the German Protestant tradition - and even those in the Catholic tradition - looks to Luther as its intellectual and theological ancestor (Lohfink, 2001). The choice to look at Hfa was not made lightly. There are other translations that are more controversial, even more dynamic. The quality that makes Hfa the choice for this project lies in the systematic nature of its translation. On the back cover of the “Trend Edition” of Hfa, below a picture of three popular looking teens surrounded by a picture of graffiti, is this description of the work: “Die Bibel, die unsere Sprache spricht” (The Bible, that speaks our language) (Hfa, back cover, translation mine). In the foreword to the Bible, the publishers state the following as the motivation for their translation:

Eine gelungene Übersetzung soll nicht nur die Botschaft des Originaltextes zuverlässig wiedergeben, sie muss auch verständlich sein, natürlich und lebendig klingen – so wie wir uns in unserer Sprache ausdrücken. Kurzum: Sie soll auf ihre Leser möglichst die gleiche Wirkung haben, wie sie das Original auf die damaligen Leser hatte!

A successful translation should not only reliably repeat the message of the original text, it must also be understandable, sound natural and alive – as we express ourselves in our language. Briefly: it should have the same effect on its readers as the original had on its readers of that time! (p. XXVIII, translation mine).

The Hfa New Testament was published with the Old Testament Book of Psalms in 1991, and a complete version with both testaments followed in 1996. Work on the Old Testament was done in cooperation with the International Bible Society (IBS) (Übersetzung des Alten Testaments, http://www.hoffnungfueralle.com accessed in 26-03-
The source text for the Old Testament was the Masoretic text of the Old Testament in Hebrew. Other translations, such as the Greek Septuagint were brought into play as well for the translation (Hfa, p. XXIX). The complete text was then revised in 2002, with the following goal: “die Qualität der Übersetzung im Blick auf Zuverlässigkeit und Verständlichkeit weiter zu verbessern” (to further improve the quality of the translation in relation to reliability and understandability) (ibid., translation mine).

Two key terms seem to form the foundation for the translation goals of Hfa – Zuverlässigkeit and Verständlichkeit – reliability and understandability. These terms have appeared already in the quote above, and they are repeated throughout the forward to Hfa. According to the publishers, these goals cannot be met by doing a word for word translation. Rather, the translators sought to take the sense (Sinn) of the Greek and Hebrew texts and express it in comprehensible German (p. XXVIII).

In order to maintain reliability and understandability, the translators followed the method of functional or dynamic equivalence, as proposed by Eugene Nida (Übersetzungsprinzipien, http://www.hoffnungfueralle.com, last accessed on 30-03-2007).

Nida (1964) differentiates between two forms of equivalence: formal and dynamic (p.159). Formal equivalence places importance on the goal that “the message in the receptor language should match as closely as possible the different elements in the source language” (ibid.). On the other hand, a dynamic translation “is not concerned with matching the receptor-language message with the source-language message, but with the dynamic relationship, that the relationship between receptor and message should be
substantially the same as that which existed between the original receptors and the message” (ibid.). According to Statham (2005), this form of translation was initially meant only to refer to “objective entities” (p. 30). That is, when an object that existed in the original Bible language did not exist in the target language, the translator was to go to the “closest functional equivalent in the culture” (ibid). Early on, Statham states, Nida’s method did not focus on linguistic entities, such as syntax and aspect (p. 31). Later, however, the theory was extended to linguistic entities, such as in the case of many sentences in the Bible beginning with “And” (p. 33). Later discussion will examine how this opening of form to change in translation becomes a controversial aspect of Hfa. The efforts of these translators have seen success in sales. According to Felber (2004), Hfa was one of the most successful new German Bible translations. The “Trend Edition”, discussed here, was already in its fifth printing by 2003, with the complete Bible having gone through 14 editions (p. 181).

Felber (2004) takes to task the methods used in translating Hfa. He states that the methods of Nida, which claim to promote loyalty to sense rather than form, do not in fact do so. Rather, this method opens the door to arbitrary additions to the Bible (p. 182). In his argument, he lends further support to the idea that humans in Hfa are given a more active position over God (p. 184). Felber also criticizes the changes of certain key terms, such as Evangelium (Gospel) to weaker forms such as Heilsbotschaft (healing message) and other forms (pp. 189-191). According to Felber, “Man sollte also nicht die “Hfa” 1983 zur Hand nehmen, wenn man das “Evangelium” klarmachen will” (One should not use “Hfa” 1983 if one wishes to make “Gospel” clear) (ibid., translation mine). Felber’s overall critique of Hfa comes down to one key issue. By changing the language, in this
case to something more flowing and journalistic, the strength of the text, as it is found in Luther, is lost (p. 199).

Haubeck (2002) adds support to the argument that there are problems with the Hfa translation. He argues that Hfa causes the text to lose much of its strength due to changes or shortenings in the sense of the text. He gives Matthew 7:3-5 as an example of this. In Luther, the verse speaks of removing the large piece of wood from your eye before trying to help your friend with the splinter in theirs. Haubeck cites a version of Hfa where the wood metaphor is simply replaced with “mistakes”, that you should deal with your own mistakes before trying to help or criticize others (p. 256). It should be noted, however, that in later editions of Hfa this imagery has been replaced.

The Evangelical Church of Germany (EKD) also provides commentary on Hfa. The EKD review of Hfa is quite striking. It is brief, but rather damning in its accusations towards Hfa:


This translation seeks to reach understanding through everyday German. It uses additions, shortenings or rewritings of the Bible text, without normally making the reader aware of these. With that come translations and explanations of the text that are wrong or contrary to the sense of the text, which lack any foundation. Where the Bible text consciously uses the same formulation, the translation is so changed that the German reader can not recognize the biblical characteristics. Despite some corrections in
the revised edition, this Bible must be used in conjunction with carefully translated versions. Hfa is unfit for use in the congregation. (http://www.ekd.de/bibel/bibel.html, last accessed on 30-03-2007, translation mine).

This critique of Hfa remains even after its revision in 2002, which claims to have dealt with many of this issues mentioned above.

It is clear that both Luther and Hfa have their strong supporters and detractors. Most of this discussion has, however, taken place within the realm of theological debate, despite the fact that much of the data in question could be considered linguistic in nature. Issues of linguistics and readability have remained more in the academic sphere. In the next section, I shall discuss more of how such areas could be brought together to broaden the discussion of the merits of differing Bible translations.

1.5 Literacy, Orality and Religiosity

As this project deals with how readable the two Bible versions were for people of different times, it is important to understand the differences between the peoples of the 16th and 20th centuries. Three aspects are important here: 1) literacy, examining the portion of the population that could read the Bible; 2) orality, examining the language people were used to and how the Bible was presented; 3) religiosity, examining what portion of the population considered themselves believers in the message of the Bible translations.

Major differences can be seen in each of the above areas. From the standpoint of literacy, according to Scribner and Dixon (2003), literacy was very low in Germany at the time of Luther, at 4-5%, or around 400,000 people out of a population of 16 million (pp.
19-20). Thus, at this point, “the spread of evangelical ideas depended as much on forms of oral communication as it did on printing” (p. 20). In contrast, Germany today, according to a 2003 CIA estimate has a 99% literacy rate for people over 15 years of age (www.cia.gov, last accessed 6-23-2008).

As previously stated, the German-speaking lands of the 16th century still contained a largely oral culture. According to Ong (1982), people in oral cultures receive and transmit information differently than those in literate cultures. Some aspects of oral thought are particularly important for this study. Oral language is additive rather than subordinative. This indicates that coordinating conjunctions are used far more often than subordinating conjunctions in oral traditions (p. 36). Oral language is also aggregative rather than analytic, relying on “formulas to implement memory” (p. 38). Finally, oral language is “redundant or copious” (p. 39), often repeating things that were just said. The quantitative data from this study has the potential to touch upon each of these areas, giving insight into the nature of the reception of the two Bible versions.

It is important to also understand who is receiving these texts, as religion has played very different roles between the 16th century and the present. During the time of Luther, Germany was in a time of religious upheaval between Catholics and followers of Luther’s Reformation, although for much of that time the movement was “vague, undefined, heterogeneous and unstructured” (Hillerbrand, 1981, p. 28). Luther’s message spread quickly for the time, with 300,000 copies of his Bible translation produced between 1517 and 1520 (Scribner and Dixon, 2003, p. 20). Coupled with the spread of the evangelical message via sermons (p. 21), the Reformation spread quickly through much of what is Germany today.
In modern Germany, however, the number those who openly declare themselves as evangelical Christians in population statistics is decreasing. According to the Statistisches Bundesamt Deutschland “German Federal Bureau of Statistics”, out of a population of approximately 82,438,000 people in 2005, just over 25.3 million people were listed as evangelical Christians, or about 30.7% of the total population (www.destatis.de, last accessed 6-23-2008). This could also have an effect on how people receive the Bible. With a low percentage of people attending evangelical churches, it is possible that those who do receive the Bible do so privately, as opposed to the people of the 16th century, who, as we saw above, did so primarily orally. One of the questions for this project is whether or not this is evident in differences in the two Bible versions.

1.6 Readability

At the core of this study is the idea of readability. According to Bruce and Rubin (1988), formulae that examine this phenomenon attempt to assign “to a text a numerical estimate of ‘readability,’ variously defined as ‘ease of reading’, ‘interest’ or ‘ease of understanding’” (Gilliland, 1972, quoted in Bruce and Rubin, 1988, p. 5). Such formulae are intended to be “quick and convenient measurement[s]” (ibid.), and thus take “into account only easily measurable aspects of text such as word difficulty and average sentence length” (pp. 5-6). It will be clear as this text progresses, however, that it is by no means a simple task to examine the readability of a text. While such aspects as sentence length or word difficulty may seem like simple objects to be quantified, the very
choice of such objects can lead to highly differing and incompatible measures of readability. Indeed, the chosen object of analysis can be considered one of the key steps in the process of studying readability. While different methodologies of studying readability will be discussed in a later chapter, this section will be devoted to looking at those researchers who study readability and why they do so.

### 1.6.1 Who Studies Readability

Readability is typically studied in relation to two groups – producers of texts and consumers of texts. People who are interested in readability come at the problem from a number of different perspectives as well. One area is in the development of reading texts for different situations, both for L1 and L2 learners. An example of this would be Charrow (1988), who examined readability formulae as guidelines for improving a text for reader comprehension. Another prominent area of interest in readability is in the realm of textbooks outside of the instruction of reading. Studies have been done on how textbook information can be difficult for novice readers in an area (Halliday, 2004), as well as looking at strategies for unpacking the information in academic texts (Ventola, 1996; Unsworth, 1999). Halliday (2004) illustrates the problem in this way:

> In any typical group of science students there will be some who find themselves in difficulty – who find the disciplines of physics, or biology, or mathematics forbidding and obscure. To such students, these subjects appear decidedly unfriendly. When their teacher tries to diagnose the problems the students are having, it is usually not long before the discussion begins to focus on language. Scientific texts are found to be difficult to read; and this is said to be because they are written in “scientific language”. A “jargon” which has the effect of making the learner feel excluded and alienated from the subject-matter. (p. 159)
This is not only the case in texts dealing with physical sciences and mathematics. Unsworth (1999) examines this issue in the realm of history texts, as well as science texts. Unsworth sees the following points as key issues in “developing an understanding of the distinctive literacy demands of school science and history texts” (p. 509):

- Distinguishing the grammar of specialized knowledge from the grammar of everyday language;
- Understanding the complete interconnections of content area learning and learning to control the grammatical forms characteristic of texts in different content areas;
- Using functional grammatical concepts to differentiate the literacies of different content areas; and
- Developing functional grammatical knowledge as a tool for critical literacy. (ibid.)

As mentioned above, those who study the complexity of texts come from a variety of viewpoints concerning the source of textual complexity. They have studied a number of textual aspects and attributes: syntactic complexity, lexical complexity, inferential complexity, word length, sentence length, lexical density, and others. Despite theoretical and methodological differences, however, most researchers in this area are working towards a common goal – the search for an empirical means of determining the difficulty of a text and from that, to develop ways to make texts more accessible to readers.

While the proper empirical measure of readability may easily be disputed, it is accessibility that is a key point for this project. An important aspect of Bible translation is making the text of this book accessible to as many people as possible. Unlike textbooks, which typically focus on one subject area, Nida (1964) demonstrates the complexity of the realm of Bible translation and the difficulties with which a translator seeking to produce a more readable version of the Bible would find him or herself:
Of all the various types of translating, however, one can safely say that none surpasses Bible translating in: (1) the range of subject matter (e.g. poetry, law, proverbs, narration, exposition, conversation); (2) linguistic variety (directly or indirectly from Greek and Hebrew into more than 1,200 other languages and dialects); (3) historical depth (from the third century B.C. to the present); (4) cultural diversity (there is no cultural area in the world which is not represented by Bible translating); (5) volume of manuscript evidence; (6) number of translators involved; (7) conflicting viewpoints; and (8) accumulation of data on principles and procedures employed. (p. 4)

Clearly, in the task of producing a text that is reliable and readable, the work of a Bible translator is extremely difficult.

1.7 Motivation for Study

The motivation for this study is a desire to take part in the long standing discussion of different Bible translations from a systematic, corpus analytic perspective of readability. As will be seen in the next chapter, up to now, the methods I use in this work have been primarily applied to academic and scientific texts. This has been done by various authors in order to ascertain how accessible the knowledge found in such texts is for students. Such methods will be applied to the text of the Bible, a book from which many have sought knowledge and understanding for centuries.

If, as Dewey states, readability is a goal of Bible translation, then the grammar of these different translations must be closely scrutinized by principled means. The hypothesis for this study is that the translators of both texts endeavored to make their Bible versions as readable as possible for their contemporaries. Readability is a desirable characteristic of any translation. One of the justifications given for both Luther and Hfa was the perceived need for a translation that was more readable for contemporary
Germans. Therefore, this study applies quantitative measures of readability to both HfA and Luther’s translations in order to assess the relative readability of the two for their respective contemporary audiences.
Chapter 2
Background and Methodology

2.1 Introduction

In this chapter I explore the various methods and theoretical constructs that will form the basis of this study. The research underpinning each of these areas and the methods that I will employ to explore them will be discussed in the subsequent sections of this chapter. In order to understand why I have chosen the methodologies that I have, it is important to recapitulate the research questions for this project.

3. Do Luther and Hfa differ in readability as assessed by a range of linguistic measures? If so, how do they differ?
   a. are there differences in lexical density?
   b. are there differences in clausal complexity?
   c. are there differences in the use of grammatical metaphor via nominalization?
   d. are there differences in the use of formulaic language in the form of lexical bundles between the two versions?

4. Are the linguistic differences identified by measures of readability driven by linguistic changes in the German lexicon and syntax since the mid-16th century or by changes in the readership and use of the Bible since that time?

In order to answer these questions, several things must be done first. The primary goal of this chapter is to link each of the four areas listed under question one to the idea of readability, and thus to the other overarching questions. While some of them have not been directly linked to readability in current literature, it is my view that they play an
important role. Secondly, it is the goal of this chapter to broadly describe how the study will proceed through the two research questions. In order to compare the readability of these two texts, four linguistic areas will be examined: 1) lexical density, 2) clausal complexity, 3) nominalization, and 4) lexical bundles. Methods of corpus analysis will be used to look at each of these phenomena.

This chapter will proceed through each of the main research questions and their sub-questions. Each section will describe the object of linguistic analysis and critically discuss various methods that have been used to examine this area. This will be followed by a discussion of the methodology to be used for this area in this study, and the rationale behind that choice. Each section will conclude with a discussion of how the area relates to the overall research questions. It is also important to note that there will be a discussion of the general means of analysis for this project – corpus analysis, and its importance to this project and its applicability to the questions at hand.

2.2 Corpus Analysis

I begin this discussion with the primary tool to be used for gathering and analyzing data – corpus analysis. In the quote below, Sinclair (2004) describes the power of corpus analysis in the following manner:

A corpus is a remarkable thing, not so much because it is a collection of language text, but because of the properties that it acquires if it is well-designed and carefully-constructed. (Sinclair, 2004, http://ahds.ac.uk/creating/guides/linguistic-corpora/chapter1.htm, accessed on 08-20-2007)
A corpus based approach to analysis is defined by Biber, et al (1999) as a method in which “grammatical descriptions are based on the patterns of structure and use found in a large collection of spoken and written texts, stored electronically, and searchable by computer” (p. 4). Note also that many of the studies to be cited in this work are corpus analytical in nature. Corpus analysis allows us to systematically examine large pieces of text to look for statistical trends, with the possibility of focusing on a variety of areas.

This work employs parallel corpus linguistics. According to Borin (2002), there are many ways of describing corpora which could be considered to be parallel (p. 2). Below are some definitions that are particularly applicable:

**Parallel corpus**

a ‘collection of translationally related texts’

“Two (or more) subcorpora which exhibit some kind of parallelism” (Ebeling 1998a, quoted in Borin, 2002, p. 2)

a “collection of functionally similar (original) texts (in two or more languages)” (Hartman 1997, quoted in Borin 2002, p. 2)


“an equal amount of texts originally written in language A and B and their respective translations” (ibid.)

“only translations of texts into the language A, B and C, whereas the texts were originally written in language Z” (ibid.)

**Diachronic Corpus**

“For example Chaucer’s Canterbury Tales in Medieval Enhlsih vs. modern English versions” (Merkel 1999, p. 11, cited in Borin 2002, p. 3)

**Target Variant Corpus**

“Different translations into the same target language of the same original text” (ibid.)
The last two definitions, in particular, play a role in this study, as the two translations under investigation have a relationship both as diachronic versions of the same text, and as different translations of the same source material. Borin cites three areas as the “most well-known uses of parallel and comparable corpora” (p. 14):

1. “for contrastive and typological grammatical and lexicographical studies in linguistics.”
2. “for knowledge acquisition for machine translation in computational linguistics”
3. “as a source of authentic contrastive language data in language learning and teaching.” (ibid.)

For the purposes of this study, I will be focusing primarily on the first area.

The use of parallel corpora lends itself well to studies in many different areas. Smith (2004), for example, used parallel corpora of advertisements in English and Russian to compare the use of personal pronouns and possessive determiners in that genre in both languages. Other studies, such as Resnik and Smith (2002) and Li and Yang (2006), examine the nature of the World Wide Web as a parallel corpus.

Corpus analysis has already been used to do linguistic studies of various versions of the Bible. Wehrmeyer (2004) examined the feasibility of corpus analysis as an aid to Bible translation. According to the analysis done in that article, corpus analysis has much to offer the Bible translator, and the Bible has much to offer the corpus analyst. The translator has a “fast and rigorous method of assessing his product” and Bible corpora offer the corpus analyst “not only a wide variety of genres but also carefully checked translations (pp. 223-224). Naude (2004) utilized corpus analysis to look at the translation of biblical poetry into Afrikaans, focusing on whether or not aspects of
parallelism are carried from the Hebrew Bible into translations in Afrikaans. The Bible has proven to be an ideal testing ground for cross-linguistic corpus linguistics. According to Resnik et al. (1999), the Bible is “a widely available, representative sample of carefully translated texts in a variety of styles in a broad range of languages” (p.129). This fits very well with the definition that Biber et al. (1998) state for what a corpus should do: “a corpus seeks to represent a language or some part of a language. The appropriate design for a corpus therefore depends upon what it is meant to represent” (p. 246). According to Sinclair (2004), several criteria must be determined in the construction of a good corpus (http://ahds.ac.uk/creating/guides/linguistic-corpora/chapter1.htm, accessed on 08-20-2007):

- the mode of the text; whether the language originates in speech or writing, or perhaps nowadays in electronic mode;
- the type of text; for example if written, whether a book, a journal, a notice or a letter;
- the domain of the text; for example whether academic or popular;
- the language or languages or language varieties of the corpus;
- the location of the texts; for example (the English of) UK or Australia;
- the date of the texts. (ibid.)

The Bible, in this case, is a representative text for written Christian thought in any language, as this book forms the basis for that belief system. The domain and language of the texts under investigation is also unified, both being from parallel biblical texts in German. Only one language is contained in the corpus, and the texts are from well established language varieties – the German of Martin Luther, and the modern German of Hfa. Finally, the dates of the development of the texts are well established, following the dates of revisions to both versions. Other aspects also make the Bible particularly suited
to corpus analysis, both within a language and cross-linguistically. The following
grounds for the Bible making a good corpus are paraphrased from Resnik et al (1999):

- The Bible is among the world’s most carefully translated texts
- Translations can be aligned with their contemporaries from other languages
- Structure is largely standardized
- Variation within verses does not affect verse alignment

While Resnik et al. view the Bible as a good corpus for cross-linguistic study, I
believe such a corpus is well suited for this project, because I am looking at variation in
this use of German over time. Thus, I assume that although people may have their
various problems with one translation or another, both texts under investigation here have
been carefully translated. The alignment of verses will prove especially important, as it
will allow for a standard annotation for the corpus, as well as for easy pinpointing of
differences. The aspects to be explored via corpus analysis will be further discussed in
the following sections of this chapter.

2.3 Readability

As stated in the introduction, this project finds its basis in the idea of readability.
It is important to have an understanding of current ideas behind readability prediction as a
place to start for further arguments. Readability is by no means a new area of research.
According to Chall (1988), studies in readability have “deep roots in the classical rhetoric
of Plato and Aristotle and in the vocabulary analyses of the Bible by ancient Hebrew
scholars” (p. 2). In more modern times, Davison and Green (1988) report that research in
this area began in the 1920’s, looking for correlations between “objectively observable
features of texts and the reading levels of readers, as measured by standardized tests” (p. 1).

Researchers have employed a number of different formulas for this purpose. Each one focuses on different aspects of a text, and each has a different, and generally non-comparable, means of assigning value to their predictions. Despite this, Davison and Green (1988) point out some commonalities between most of them: “The text properties usually are average sentence length, normally based on samples of 100 words, and an estimate of word-difficulty, typically based on syllable length or occurrence on a list of high-frequency words” (p. 2). The authors argue, though, that such formulas ignore factors of the text and the reader that could be of great importance.

According to Davison and Green (1988), studies of readability have had two sorts of foci. The first has to do with understanding from the perspective of the hearer or reader – that is, “how do people identify and interpret linguistic forms, how do they derive information from a linguistically encoded message, and how do they integrate new information with previously known information?” (p. 4). The second question deals with the objective features of the text itself. The authors ask, “Do features of the text reflect its difficulty, and do linguistic features – such as word difficulty and sentence complexity in themselves present barriers to comprehension?” (ibid.).

2.3.1 Readability Formulas

This section of the chapter will examine several different metrics for readability which are well known and remain in use today. I will present the basics of these
measures, along with a critical discussion of what they share in common, what sets them apart from one another, and what their limitations are for a study such as the one in this work.

**Flesch Reading Ease**

The first metric to be examined here is the Flesch Reading Ease formula (Flesch, 1949). This early formula still finds some use today, and aspects of it can be seen in other measures, as will be seen below. Instructions for applying the formula to a text are simple, as can be seen below:

- Pick your samples
- Count the number of words
- Figure the average sentence length
- Count the syllables
- Count the “personal words”
- Count the “personal sentences”
- Find your “reading ease” score
- Find your “human interest” score (Flesch, 1949, pp. 247-251)

Once the user of this formula finds the prescribed scores from the directions above, Flesch provided a special chart for determining the ease of reading of a text, as can be seen in figure 2.1. Factors in this model include a word count (typically sample sizes of up to 100 words), average length of sentences and the number of syllables contained in the entire passage (pp.247-248). “Personal Words” include all pronouns except those referring to things, all words with masculine or feminine “natural gender” and all words referring to groups of people (pp. 248-249). Having calculated all of these values, one could either use the table below or the following formula:
Multiply average sentence length by 1.015
Multiply number of syllables per 100 words by .864
Add
Subtract this sum from 206.835
Your “reading ease” score is …… (p. 250)

Figure 2.1: Flesch Reading Ease Chart  (Flesch 1949, inside cover)

In this model, complexity is defined solely by sentence length and syllable counts. As I show other methods, it will be clear that some aspects of this measure are maintained, while others are added.
Fry Graph for Estimating Readability

According to Klare (1988), the Fry Graph is “the most widely used of all readability methods”, which, at the time, was attested to by the existence of a hand calculator specifically for the graph, as well as a parallel computer program (p. 22).

Figure 2.2

Figure 2.2: Fry Graph (http://school.discovery.com/schrockguide/fry/fry2.html accessed 8-27-2007)

Reading ages are established with this method in the following way:

- Randomly select 3 one hundred word passages from a book or an article
- Plot average number of syllables and average number of sentences per 100 words to determine grade level
- If variability is observed, conclude that readability is uneven
- Count proper nouns, numerals, and initializations as words – one syllable per symbol.
As with the Flesch model, the Fry graph focuses primarily on the number of syllables and average sentence length. Rather than simply providing a number for ease or difficulty, this model correlates the results with an age or grade level for readers, providing more guidance for teachers and pedagogues using this model.

**Gunning FOG Index**

The Gunning FOG Index is cited by Ventola (1996) as a measure of readability. Ventola quotes Ellis and Hopkins and their presentation of this measure, as it is done in a very clear and concise manner (p. 157). It functions in this way:

- take a sample passage of about 100 words
- count the number of words in the sentences
- divide the number of words in the passage by the number of sentences, giving the average length of sentence in the passage.
- count the number of words of three or more syllables in the passage (not counting proper names, compound names…)
- total the two factors just counted and multiply by 0.4 to obtain the Fog Index (Ellis and Hopkins 1985, pp. 35-36, cited in Ventola, 1996, p. 157).

According to this method, a score of 13 or more is considered a difficult text (pp. 157-168). Ventola employs the FOG index in concert with lexical density, a measure of readability which will be discussed more in depth later, in order to examine how information is packed and unpacked in academic texts.

**Dale-Chall Readability Formula**

A similar method to the FOG index is the the New Dale-Chall Readability Formula (Dale and Chall, 1995, cited in Kotula, 2003). This method uses similar sample sizes, but also focuses on aspects like word frequency, as can be seen in this description of the process:
• Count an exact 100-word sample from an early part of the book.
• Count the number of complete sentences.
• Note the number of words in the sample that do not appear on the Dale list of 3,000 words known by students in grade four.
• Find the cloze score and reading level from the tables provided.
• Repeat this process for every 50 pages of text.
• Average the results of the samples to compute the average cloze score and reading level. (Kotula, 2003, p. 195)

According to Kotula, this method also uses worksheets to examine aspects such as characteristics of the reader and “cognitive-structural aspects of the text” (ibid.).

### 2.3.2 Criticism of Readability Formulas

According to some authors, attempts to quantify and predict difficulty have had some challenges, especially in terms of whether or not the objects of study actually contribute to reading ease or difficulty. Bruce and Rubin (1988) state that “readability formulas fail in many of their applications” (p. 7). They cite two factors for this:

1. Formulas contradict what is currently known about reading and the reading process.
2. Such formulae have been used in manners that make the situation worse (ibid.).

Further related problems are cited by Harrison and Bakker (1998, p. 124):

1. *Lack of clarity about purpose*: formulae are often used as guidelines for revision rather than their intended use as predictors.
2. *The questionable validity of prescribing short sentences*: such sentences are often prescribed despite research that combining sentences could improve comprehensibility.
3. *Readability formulae ignore the reader, reading style, and context*: many formulae do not take any of this information into account.
4. *Lack of clarity about the meaning of readability scores*: it is unclear what reading level scores imply, and even what is meant by understanding.
To counter this, the Harrison and Bakker looked at a more systemic functional approach to readability, focusing on lexical density, which will be discussed in the next section.

2.4 Lexical Density

Lexical density is my chosen method for initial readability prediction. According to Halliday and Matthiessen (2004), lexical density is the primary measure of complexity for written texts (p. 654). Ure (1971) defined lexical density as the proportion of lexical items to total orthographic words. Essentially, lexical density could be counted in this method using the following formula:

\[
\text{Number of lexical items} \times \frac{100}{\text{Total number of words}} = \text{Lexical Density}
\]

(Ure, 1971, p. 445)

Essentially, in this method, lexical density is the percentage of lexical terms in the total count of orthographic words. In Ure’s observations, spoken texts were typically under 40%, and written were over (Ventola, 1996, pp. 158-159).

There is an important feature from the definition that sets lexical density apart from other readability predictors, namely the focus on the functions of the words in the text – lexical words versus grammatical words, as opposed to aspects such as sentence length and syllable counts. Lexical words are defined as content words – being names of entities, actions, and qualities (p. 37). Included here are nouns, verbs, adjectives and adverbs. Other types of words function more grammatically, that is, they do not carry
Halliday (1989) differentiates between lexical and grammatical items in the following way:

A grammatical item enters into a closed system. For example, the personal pronoun *him* contrasts on one dimension with *he, his*; on another dimension with *me, you her, it, us, them, one*; but that is all. There are no more items in these classes and we cannot add any. With a lexical item, however, we cannot close off its class membership; it enters into an open set, which is indefinitely extendable. So the word *door* is in contrast with *gate, and screen*; also with *window, wall, floor and ceiling*; with *knob, handle, panel, and sill*; with *room, house, hall*; with *entrance, opening, portal* – there is no way of closing off the sets of items that it is related to, and new items can always come into the picture. (p. 63)

In a comparison with other measures of readability, Harrison and Bakker (1998) found lexical density to be superior to other methods in gauging the readability of similar texts. This study compared lexical density with other commercially available methods of rating the readability of texts. Similar pairs of texts were rated for readability using the different methods (p. 126). Subjects were then asked to choose which was easier to read (p. 127). Their preliminary findings in this pilot study were that lexical density proved to be a more accurate predictor of what the subjects would choose than the other methods used (p. 131).

Further studies of lexical density have focused on lexical density in primarily academic texts. Halliday (2004) examines problems in the grammar of scientific texts partially from the perspective of lexical density, making particularly harsh critique of scientific writers who make texts unnecessarily complex (pp 178-179). Unsworth (1989) also utilized lexical density as part of an examination of scientific and historical texts, in the realm of student literacy. Ventola (1996) used lexical density in combination with the FOG index mentioned above to examine academic writing and the ways in which writers
pack and unpack information. In each of these studies, lexical density was used as a factor in text difficulty, along with other aspects, such as grammatical metaphor and grammatical complexity, which shall also be discussed in this project in the form of nominalization and clausal complexity, respectively.

In this study, lexical density will be calculated using the formula seen above from Ure (1971). This will be accomplished using part-of-speech tagging, which will also be used to calculate the concentrations of nouns and verbs within the set of content words in the two corpora.

### 2.5 Clausal Complexity

The second area of exploration is the clause. According to Halliday and Matthiessen (2004), “[t]he clause is the central processing unit in the lexicogrammar – in the specific sense that it is in the clause that meanings of different kinds are mapped into an integrated grammatical structure” (p. 10). The authors differentiate between three different types of clausal relations here – parataxis (independent), hypotaxis (dependent), and rankshift (embedded). For the purposes of this study, the focus will be on markers of parataxis and hypotaxis, namely coordinating and subordinating conjunctions.

Halliday states that “[p]arataxis is the linking of elements of equal status” (p. 384). This relation is normally symmetrical and transitive (ibid.). For our purposes, it is exemplified in two independent clauses joined by “and” or another coordinating conjunction, such as “then”, as seen in the following example:

Table 2.1
Table 2.1: Parataxis (Halliday and Matthiessen, 2004, p. 387)

<table>
<thead>
<tr>
<th>I went to school in New York City</th>
<th>and then we lived up on the Hudson for a while,</th>
<th>then moved to Connecticut.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>x2</td>
<td>x3</td>
</tr>
</tbody>
</table>

In the sentence above, the clause complex is expanded using *and* and *then*.

Hypotaxis denotes “the binding of elements of unequal status” (p. 37), such as a dependent clause starting with “that” or “when”, as seen in the next table:

Table 2.2

<table>
<thead>
<tr>
<th>When all had been done</th>
<th>as God had ordered</th>
<th>Noah closed the door</th>
</tr>
</thead>
<tbody>
<tr>
<td>xβ</td>
<td>xγ</td>
<td>α</td>
</tr>
</tbody>
</table>

The third clause in the sentence above, marked alpha, is so marked because it is the independent clause of the sentence. The first clause is bound directly to the independent clause, thus being marked beta, and finally clause two is bound to the beta clause.

While the computation of lexical density is completed independently of the clause, clausal structure is clearly important for readability. The importance of looking at lexical density in relationship to clauses is adapted from Halliday well by Ghadessy and Gao (2001):

Excluding any consideration of clauses of any text in studies of word frequencies neglects a very important observation by Halliday that the clause “is perhaps the most fundamental category in the whole of linguistics.” The building blocks of texts are not words or phrases by clauses which function as (a) “the representation of phenomena of experience,” (b) “the expression of the speech function,” and (c) “the bearer of the message”. Halliday’s formula for measurement of lexical density, i.e. total number of content/lexical words divided by total number
of ranking clauses, gives due credit to the significance of this linguistic unit in analyzing text. (p. 68, adapted from Halliday, 1985, p. 67).

Clausal complexity is important for readability because, at its core, parataxis and hypotaxis are matters of grammatical intricacy. According to Halliday (1989, p. 87), in spoken language, “Grammatical intricacy takes the place of lexical density”. Therefore, as clausal complexity increases, lexical density could decrease, as the text becomes more like spoken language. Clausal relations have been looked at in other areas of readability as well. Kemper (1988) examined clausal relations from the standpoint of the comprehension of event chains. This model states that “the overall difficulty of a text should reflect how easily readers can establish the causally connected event chain underlying the text” (p. 149). Other research, such as that of Crain and Shankweiler (1988) look at individual aspects of clausal relations, such as the syntactic complexity of relative clauses.

In this study, I examine the concentration of conjunctions as a representation of clausal complexity, looking for differences in frequency of both coordinating and subordinating conjunctions, and the statistical relationship of those frequencies to lexical density.

2.6 Nominalization

When examining the differences in readability between the two texts, it is important to look not just at the statistics of lexical density, but the types of lexical items that contribute to it. Nominalization is a key term in this area, as will be seen below.
According to Halliday and Matthiessen (2004, p. 636), nominalization is a form of grammatical metaphor, which makes a text more challenging, especially for younger readers. In such nominalizations, “processes and qualities are construed as if they were entities” (p. 637). According to Taverniers (2003), written language has more such metaphors than spoken language, which is “attributed to a more general difference in types of complexity: written language is said to be ‘lexically dense’, whereas spoken language is ‘grammatically intricate’” (p. 9), thus returning us to the discussion of readability from above.

Much work has been done on nominalization in the realm of scientific writing. Banks (2003) states a number of working hypotheses on the development of nominalization in English scientific texts:

1. There has been an increasing use of nominalized processes in scientific discourse over the last 250 years.
2. Initially the physical sciences developed more rapidly than the biological sciences in this respect.
3. The difference between the physical and the biological sciences disappeared by the beginning of the twentieth century, that is about the time when the biological sciences became experimental as opposed to purely descriptive.
4. During the course of the twentieth century there has been an extension of the use of nominalized processes from the function of head in the noun group to that of modifier. (pp. 143-144)

Such developments, as can be seen in Unsworth (1999), have led to difficulties in the processing of scientific and even historical texts by students. According to Unsworth, “For students who are confident users of spoken English but lack familiarity with the grammar of the written form, the greater lexical density of school texts can contribute to comprehension difficulties” (p. 510). The increase in the use of nominalizations in such texts is undoubtedly a contributing factor. The ability to construct and process the
specialized knowledge of scientific texts is dependent on the reader’s ability to process such forms (p. 511). According to Halliday (2004), some authors, especially in scientific texts, make this difficult. He gives this harsh criticism:

…it must be said that many of those who write in the language of science write it very badly. They leave implicit things that need to be made explicit, create multiple ambiguities that cannot readily be resolved, and use grammatical metaphor both inappropriately and to excess. The language thus becomes a form of ritual, a way of claiming status and turning science into the prerogative of an elite (pp. 178-179).

While it is unlikely that such complex language would be found in a Bible translation, such a statement does strongly underline the power of forms like nominalization and other grammatical metaphor. It is with this in mind that I pursue the matter of nominalization in Luther and Hfa. Below I will show that there are definite parallels in the development of nominalization in German with that of English.

Many sources have documented the development of nominalization in German, as well as issues involved in the use of nominalization. Burkhardt (1998) categorizes German as a nominalizing language (p.118). Two areas in particular are highly prone to greater use of nominalization, according to the author:


Next to the language of bureaucracy, science has above all an undeniable tendency towards nominalization. In place of the simple sentence *it is raining*, a well-versed use of scientific language would perhaps say: *As a result of an increase of the low-pressure-influence from the west, the current observable weather-situation is characterized, through increased*
cloud-creation, by a clear downward-falling-tendency. (ibid., translation mine).

While the example above is somewhat extreme, the point is well taken – the use of nominalization allows the writer to take a rather banal statement – *es regnet* – and make it much more complex and scientific sounding, through the use of such nominalizations as *Zunahme* “increase”, *Tiefdruckeinfluss* “low-pressure-influence”, and *Wolkenbildung* “cloud-creation”.

Waterman (1966) sees this tendency towards nominalization as something that could, one day, result in a more concrete structural alteration of German (p. 169). Although, like Burkhardt (1998) above, Waterman points to the increased wordiness of nominalization use, he does point to one advantage “of anticipating the message of the normally last-place verb, thus offsetting to some extent what has long been considered the clumsiest feature of German word order” (ibid.). He does, however, point out a complaint by some that a more noun-rich style of language leads to less dynamic and more static language (ibid.).

There are, however, differences in the views of what a trend towards nominalization could look like. Stedje (1989) gives a clear visual example of the difference between a more noun-based (Nominalstil) and a more verb-based (Verbalstil) language:

Table 2.3
Table 2.3: Nominalstil vs. Verbaler Stil (Stedje, 1989, p. 181)

<table>
<thead>
<tr>
<th>Nominalstil</th>
<th>Verbaler Stil</th>
</tr>
</thead>
<tbody>
<tr>
<td>Die Obigen Darlegungen <em>wollen Anstoß sein</em> zu einem verstärkten Bemühen</td>
<td>Was oben <em>dargelegt wurden, soll</em> einen Anstoß geben. Wir <em>wollen</em> uns stärker als bisher darum <em>bemühen,</em></td>
</tr>
<tr>
<td>um eine Effektivierung des Sprachunterrichts</td>
<td>den Sprachunterricht effektiver zu <em>gestalten.</em></td>
</tr>
<tr>
<td>im Bereich rezeptiven, berufs- und damit auch gesellschaftsrelevanten</td>
<td>so daß der Schüler die Fremdsprache besser <em>aufnehmen kann</em> und die Fähigkeiten <em>erwirbt,</em> die für seinen Beruf und damit für die Gesellschaft wichtig <em>sind.</em> (45 Wörter, 11 Verben)</td>
</tr>
<tr>
<td>Sprachkönnens (24 Wörter, 2 Verben)</td>
<td></td>
</tr>
</tbody>
</table>

The numbers show a stark contrast to the statements of Burkhardt (1998) and Waterman (1966). Rather than being wordier, the *Nominalstil* shows itself to be, in fact, more concise. That being said, the *Nominalstil* text is also far denser with regard to percentage of lexical items.

Russ (2004) examined nominalization throughout all of Luther’s writings, and found the following ranking:

Table 2.4
Table 2.4: Comparison of Top Seven Suffixes According to Type and Token (Russ, 2004, p. 252)

<table>
<thead>
<tr>
<th>Position</th>
<th>Frequency of Type</th>
<th>Frequency of Token</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-er</td>
<td>-Ø</td>
</tr>
<tr>
<td>2</td>
<td>-en</td>
<td>-e</td>
</tr>
<tr>
<td>3</td>
<td>-ung</td>
<td>-er</td>
</tr>
<tr>
<td>4</td>
<td>-Ø</td>
<td>ge+e/-Ø</td>
</tr>
<tr>
<td>5</td>
<td>-heit</td>
<td>-heit</td>
</tr>
<tr>
<td>6</td>
<td>-e</td>
<td>-en</td>
</tr>
<tr>
<td>7</td>
<td>ge+e/Ø</td>
<td>-ung</td>
</tr>
</tbody>
</table>

As can be seen from this chart, German has some clear markers of nominalization, and these are just a selection of what is possible. Russ states that “Luther’s usage is similar to Modern German” (p. 261).

The examination of nominalization in the present study will consist of two parts. First, a corpus analysis will be utilized to document the most frequent occurrences of different types of nominalizations. As is clear in the table from Russ (2004) a convenient means of categorizing nominalizations in German is by their suffixes. Once I have made clear what types of nominalizations are common in both texts, as well as which text uses more, the second stage will be to examine their use in context. Here the question becomes, when such difficult terms are present within the Bible texts, how, if at all, are they unpacked within the text itself? In order to examine this question, occurrences of one type of nominalization will be analyzed in order to ascertain if and how the text clarifies those terms.
2.7 Lexical Bundles

The final type of corpus analysis in this project will be to look at the use of lexical bundles, a type of formulaic language, in the two Bible versions. Up until this point, the study has focused primarily on the appearance and frequency of individual words, be it as lexical words or grammatical words, markers of parataxis or hypotaxis, or nominalizations. The aim of this section will be to discuss methodologies for examining groups of lexemes that recur often within the texts. According to Halliday and Matthiessen (2004), such commonly occurring collocations function almost like single lexical items (p. 577). It is conceivable, then, that such collocations could have an impact on the readability of a text. Such groupings could, hypothetically, serve to reduce the lexical density of a give passage by reducing the perceived number of lexical items.

Formulaic language has also played a role in the discussion of Bible translations. According to various criticisms of Hfa, changes in the style of the language from that of Luther have caused passages to lose their former strength (Felber, 2004; Haubeck, 2004). The Evangelical Church of Germany gave a particular critique which has importance here, saying that where the Luther Bible had purposefully used a particular phrasing repeatedly, Hfa omitted such phrases, resulting in a loss of the biblical style ((http://www.ekd.de/bibel/bibel.html, last accessed on 30-03-2007). It is the aim of this section of the project to discuss possible rolls such language could play in the two texts, and how those roles could affect the readability of the two texts.

Lexical bundles are a type of formulaic language which is identified based on frequency of recurrence. Biber, et al (1999) define lexical bundles as follows:
Lexical bundles are identified empirically, as the combinations of words that in fact recur most commonly in a given register. Three-word bundles can be considered as a kind of extended collocational association and they are thus extremely common. On the other hand, four-word, five-word and six-word bundles are more phrasal in nature and correspondingly less common. In conversation there are also recurrent two-word contracted bundles, which are composed of three grammatical word forms (e.g. she didn’t → she did not). In typical written prose, these would be expressed as three separate words; thus, these two-word contracted sequences in conversation might also be compared to three-word bundles in academic prose. (p. 992)

According to Biber, et al. (1999), in conversation, 70% of words appear in non-recurrent phrases, while 25% appear in 3-word bundles and 3% in 4-word bundles. The remaining 2% appear as phrasal or prepositional verbs. In academic writing, 79% of words appear in non-recurring phrases, and 18% in 3-word bundles, 2% in 4-word and 1% in phrasal/prepositional verbs (pp. 993-994). Highly frequent small bundles may, in turn, be part of highly frequent larger bundles. In this way, I can statistically see what sorts of word groupings are regularly used in a text or group of texts.

Most of the studies done on lexical bundles so far have been exploratory in nature. That is, they explore what sorts of lexical bundles are found in different registers, such as academic speaking and writing. These studies tabulated the most frequent lexical bundles in different genres, and then categorized them by their grammatical function. Similar to the studies by Biber et al. (1999) of conversation and academic language, Cortez (2002, 2004) has examined several different areas of student and other types of academic writing. One example of this is her (2002) look at lexical bundles in freshman composition courses. The goal of this study was to compare the most common 4-word lexical bundles found in freshmen compositions with those found in academic prose and conversation in general (p. 131). This study showed that, while occasionally using forms
that did not match with academic prose, freshmen composition students did strive to utilize forms found more prevalently in academic prose and not in conversation. That is, they avoided such conversational forms as contractions and the use of the *I* pronoun “followed by a verb of perception (e.g. *I don’t think so, I don’t know what, I think it is, I know what you*) (pp. 142-143).

Lexical bundles have also been examined in the realm of historical linguistics. Culpeper and Kytö (2002) examined lexical bundles in speech-related language of the Early Modern English period, looking at dramas and trial proceedings. Each text type had an early data set corresponding to the mid-16\(^{th}\) and early 17\(^{th}\) centuries, and a late data set, corresponding to the late 17\(^{th}\) to the mid-18\(^{th}\) centuries (p. 50). This afforded the researchers both synchronic views of the specific time periods and their lexical bundles, as well as a diachronic look at the development of lexical bundles over time within each genre. One instance of diachronic change noted in this study was a trend towards more conversation-based lexical bundles – utterance launchers – in the trial data (p. 60). Granted, this can sometimes be difficult, as, more often than not, such bundles do not form any sort of complete phrase or clause, and cannot be replaced by single word/phrase synonyms.

It could be questioned here why this study focuses on lexical bundles and not other formulaic sequences like idioms. While it may be the case that idioms are very important and more likely to stored as single units (considering that the meaning of the whole typically differs from the means of the constituent parts), idioms are often not very frequent in a particular genre, and require more intuitive and less systematic means of pursuit in the realm of corpus analysis, at least for the purposes of this project. Lexical
bundles, on the other hand, are readily available from a simple frequency search of the corpus.

These bundles, in my view, have an important connection to readability. Readability measures, including lexical density, focus on individual lexemes, sometimes within structures like clauses, sometimes not. But, to my knowledge, no work has been done in readability that focuses on lexical bundles or other formulaic sequences. It is conceivable that highly frequent bundles would aid in the reading of the text, both by being processed as one unit and by serving as cohesive markers for the reader.

This study will examine lexical bundles of differing lengths – three, four, five, and six word bundles. As stated above, much work with lexical bundles has been done in the realm of academic writing. Biber, et al. (1999) presented the following findings for academic writing:

- lexical bundles are extremely common in both conversation and academic prose
- on average, lexical bundles in conversation and academic prose occur with about the same frequency
- three word bundles are more frequent (slightly) than four word bundles
- only a few lexical bundles appear with high frequency (p. 994)

The first goal in the present study will be to simply compare the number of types of lexical bundles that appear more than 5 times in each text, as well as the total number of bundles used in that grouping. Second, I will compare the different types of bundles used to see if there is a difference in their function in the text. Once there is an understanding of the different forms present in the text and their roles, this knowledge can be further applied to the discussion of readability, looking at the following questions: What sorts of roles do such bundles play? Do they guide the reader or serve some other purpose? Is it
important to have more guidance in the text, especially for a text as important as the Bible?

2.8 Conclusion

In this chapter I have examined a number of methodological approaches to the issue of readability and related linguistic features of texts. While each alone would not give a complete picture of the complexity of a text, together, these methods can be employed to gain a more complete view of the comparative readability levels of Luther and Hfa. Each chapter has two components. The first is an examination of the statistical differences between the two texts. The second is a discussion of these differences with respect to their effects on readability.
Chapter 3
Lexical Density

3.1 Introduction

Lexical Density is the primary measure of readability for this project. As stated in section 2.4, it is a measure of readability based on the proportion of lexical words to total words in a text. Lexical density is calculated here at several levels: that of the individual verse, chapters, books, and for the entire corpus. This portion of the study also examines the types of lexical words found in the corpus, comparing specifically the frequency of nouns and verbs as a percentage of the total number of lexical words at each level. In this way, the study examines both the statistical lexical density of the different text portions, as well as what sorts of lexical words contribute to this density. There are two primary research questions here: 1) which text has the lower lexical density, and, therefore, better readability; and 2) which text displays a more nominal or verbal style, based on the concentration of nouns vs. verbs?

3.2 Methodology

Four primary tasks were involved in this portion of the study: 1) analyzing the corpus based on passages of different lengths; 2) tagging the POS of each portion of the corpus; 3) calculating lexical density based on the part of speech tagging; 4) examining the text for concentrations of nouns and verbs as part of the total number of lexical
words. The following sections will describe how each stage of the study was accomplished.

3.2.1 Corpus

The corpus for this study, as described in Chapter 2, contains the first five books of the New Testament from Luther and Hfa, including the books of Matthew, Mark, Luke, John and Acts. These were chosen due to their narrative style, as compared to later texts in the New Testament which are more epistolary in nature. Due to the parallel nature of the corpus, with each text containing the same book, chapter and verse divisions, it was clear that these divisions should be used as length boundaries for the calculation of lexical density. Beginning from the verse level, ten parallel verses were chosen at random from the texts, two from each book. At the chapter level, five parallel chapters were chosen, one from each book. Then, each book would be compared, followed finally by a comparison based on the entire corpus. In this way, lexical density statistics can be compared over passages of increasing length, in order to determine whether consistent patterns occur across the spectrum of lengths.

3.2.2 Part of Speech Tagging

Part of Speech (hereafter POS) tagging of the corpus was accomplished using the program TreeTagger (Schmid 1994, 1995) for UNIX. The tagging is based on the Stuttgart-Tübingen Tagset (hereafter STTS) for German (Schiller, Teufel and Stöckert,
A complete listing of the tags used for this chapter can be found in Appendix A. For each token in the corpus, the data from this tagging contain three pieces of information: the token, the part of speech, and the lemma of the token. Tagging was done separately for each level of comparison (book, chapter, verse).

### 3.2.3 Lexical Density Calculation

Lexical density is defined here as proportion of lexical words to total words in the text (Ure, 1971; Stubbs, 1998). The value is presented as a percentage, using the following formula, where $L =$ lexical words and $N =$ number of words in the text:

$$\text{lexical density} = 100 \times \frac{L}{N} \quad \text{(Stubbs, 1996, p. 72)}$$

According to Stubbs, the word types are divided as follows:

- Lexical words: noun, adjective, adverb, main verb
- Grammatical words: auxiliary verb, modal verb, pronoun, preposition, determiner, conjunctions (ibid.)

The various tags from STTS were grouped into lexical words and grammatical words. Due to the nature of the tag set and the tagging program, however, certain verbs, such as *haben* (to have), *sein* (to be), and *werden* (to become) were grouped into the auxiliary verb category, even though they can also appear as main verbs. Without manual disambiguation of a great number of occurrences (Luther: 1,458 occurrences of *haben*, 2,735 occurrences of *sein*, 1,350 occurrences of *werden*; Hfa: 1,699 occurrences of *haben*, 2,201 occurrences of *sein*, 1,586 occurrences of *werden*), it is impossible to separate main-verb forms of *haben* and *sein* from auxiliary forms with the POS tagging methods used here. Modal verbs have a separate category, although it is also possible for
some to appear as main verbs. Using the programming language Python, I developed a program to go through the list of tagged tokens in each text and determine the proportion of lexical words to grammatical words. In order to address the issue of auxiliary and modal verbs, the program was done in two ways, the first including all verb types regardless of whether they were tagged as lexical, auxiliary or modal, and the second excluding all verbs tagged as auxiliary and modal verbs. The programs can be found in Appendix B. The program output consists of the total number of tokens in the text, the total number of lexical words, the lexical density, and the breakdown of different categories of lexical words, including nouns, adjectives, cardinal numbers, verbs, and adverbs.

3.2.4 Noun and Verb Concentration

Based on the POS tagging done to the corpus, it is possible not only to calculate lexical density based on the lexical word/grammatical word dichotomy, but also to determine the statistical breakdown of the different types of lexical words. This is done primarily with the frequency of nouns and verbs at the various levels of analysis. The different categories for nouns and verbs found in STTS are condensed simply into nouns and verbs, regardless of form. The concentration of nouns and verbs at the levels of verse, chapter, book and the complete corpus are presented as the percentage with which each part of speech comprises the total number of lexical words. As lexical density in this chapter was also calculated as a percentage, the concentration values will be compared via correlation analysis with their corresponding lexical density values. As
with lexical density, these values have been calculated both with auxiliary and modal verbs included and without them.

3.3 Data Analysis

As stated above, the data analysis takes place here in two phases, the first examining lexical density and the second looking at the concentration of nouns and verbs within the total number of lexical words at each level of analysis. At each level, the data are presented both in tabular format and as a bar graph, in order to illustrate the comparisons that are taking place between Luther and Hfa.

3.3.1 Lexical Density

As stated above, lexical density measurements for this project began with two verses chosen at random from each of the five books in the corpus. Below is an example of two parallel verses, with lexical words highlighted in bold:

Matthew 8:13 (Luther): Und *Jesus sprach* zu dem *Hauptmann*: Geh hin; dir *geschehe*, wie du *geglaubt* hast. Und sein *Knecht* wurde *gesund* zu derselben *Stunde*.

Matthew 8:13 (Hfa): Dann *sagte* Jesus zu dem *Hauptmann*: "Geh wieder nach Hause! Was du *geglaubt hast, ist Wirklichkeit geworden."
Zur selben *Zeit* wurde der *Diener* *gesund*.

In the example from Luther, there are 22 total words in the text, 11 of which are lexical words, giving a lexical density of 50.0, with modals and auxiliaries included. The same verse from Hfa 24 words, 16 of which are lexical words, for a lexical density of 66.7.
The table 3.1 shows the lexical densities of the ten verses examined in this portion of the study:

**Table 3.1**

<table>
<thead>
<tr>
<th></th>
<th>Lexical Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luther</td>
<td>Hfa</td>
</tr>
<tr>
<td>Matthew 8:13</td>
<td>50.0</td>
</tr>
<tr>
<td>Matthew 15:9</td>
<td>42.9</td>
</tr>
<tr>
<td>Mark 7:11</td>
<td>47.6</td>
</tr>
<tr>
<td>Mark 14:21</td>
<td>58.1</td>
</tr>
<tr>
<td>Luke 4:6</td>
<td>38.5</td>
</tr>
<tr>
<td>Luke 18:26</td>
<td>70.0</td>
</tr>
<tr>
<td>John 6:13</td>
<td>60.0</td>
</tr>
<tr>
<td>John 13:5</td>
<td>48.0</td>
</tr>
<tr>
<td>Acts 11:23</td>
<td>51.9</td>
</tr>
<tr>
<td>Acts 16:39</td>
<td>33.3</td>
</tr>
</tbody>
</table>

In each example, except for the example of John 13:5, the Luther verses have lower lexical density than their Hfa counterparts. This can be more clearly visualized in the figure 3.1:

**Figure 3.1**
The total length of the text samples examined here are 214 words for the Luther verses and 220 words for the Hfa verses. Average lexical density for all verses is 50.0 (SD=10.8) for Luther and 58.0 (SD=10.7) for Hfa. The differences in lexical density at this level are significant (t(9)=-3.85, p <.05), using a two-tailed t-test.

When auxiliary and modal verbs are excluded from the calculations, we see a similar picture. As expected, due to the subtraction of words from the lexical word group, lexical density generally decreases, as can be seen in the following table:

Table 3.2
Table 3.2: Lexical Density per Verse without Modals and Auxiliaries

<table>
<thead>
<tr>
<th></th>
<th>lexical density</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Luther</td>
<td>Hfa</td>
</tr>
<tr>
<td>Matthew 8:13</td>
<td>40.9</td>
<td>50.0</td>
</tr>
<tr>
<td>Matt 15:9</td>
<td>35.7</td>
<td>53.8</td>
</tr>
<tr>
<td>Mark 7:11</td>
<td>38.1</td>
<td>43.3</td>
</tr>
<tr>
<td>Mark 14:21</td>
<td>48.4</td>
<td>47.8</td>
</tr>
<tr>
<td>Luke 4:6</td>
<td>26.9</td>
<td>33.3</td>
</tr>
<tr>
<td>Luke 18:26</td>
<td>50.0</td>
<td>54.5</td>
</tr>
<tr>
<td>John 6:13</td>
<td>55.0</td>
<td>63.2</td>
</tr>
<tr>
<td>John 13:5</td>
<td>44.0</td>
<td>47.4</td>
</tr>
<tr>
<td>Acts 11:23</td>
<td>44.4</td>
<td>53.6</td>
</tr>
<tr>
<td>Acts 16:39</td>
<td>33.3</td>
<td>47.8</td>
</tr>
</tbody>
</table>

As above, in the majority of cases, lexical density remains lower in Luther than in Hfa, as shown below in figure 3.2:

Figure 3.2
While lower, we see again that the average lexical density for Luther is considerably lower at 41.7 (SD=8.4), compared to 49.5 (SD=7.9) in Hfa. The differences in lexical density at this level are again significant (t(9)=-4.56 p <.05), using a two-tailed t-test.

The pattern continues in this way when length is expanded to that of chapters. In this case, one chapter was chosen at random from each of the five books. Once again, the statistical analysis shows that Luther is consistently less dense than Hfa:

Table 3.3

<table>
<thead>
<tr>
<th></th>
<th>Lexical Density</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Luther</td>
</tr>
<tr>
<td>Matthew 4</td>
<td>51.2</td>
</tr>
<tr>
<td>Mark 6</td>
<td>45.0</td>
</tr>
<tr>
<td>Luke 15</td>
<td>47.6</td>
</tr>
<tr>
<td>John 1</td>
<td>49.0</td>
</tr>
<tr>
<td>Acts 11</td>
<td>51.6</td>
</tr>
</tbody>
</table>

In this case, as can be seen in the figure below, there are no exceptions to the pattern where Luther is less dense than Hfa:

Figure 3.3
Figure 3.3: Lexical Density per Chapter

For this sample, the total length for the Luther texts was 4,043 words, and the total length of the Hfa texts was 4,111 words. Average lexical densities were 48.9 (SD=2.7) and 53.9 (SD=1.9) respectively. The differences were once again significant (t(4)=-9.29, p <.05), using a two-tailed t-test.

When the modal and auxiliary verbs were removed from the calculations at the chapter level, we once again see a pattern in which Luther is less dense than Hfa: Table 3.4
Table 3.4: Lexical Density per Chapter without Modals and Auxiliaries

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Luther</th>
<th>Hfa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matthew 4</td>
<td>48.2</td>
<td>52.3</td>
</tr>
<tr>
<td>Mark 6</td>
<td>40.4</td>
<td>45.9</td>
</tr>
<tr>
<td>Luke 15</td>
<td>39.9</td>
<td>44.8</td>
</tr>
<tr>
<td>John 1</td>
<td>41.3</td>
<td>46.9</td>
</tr>
<tr>
<td>Acts 11</td>
<td>47.0</td>
<td>49.6</td>
</tr>
</tbody>
</table>

Even with the two grammatical verb types excluded, Luther is still less dense, often very much so, as can be seen below:

Figure 3.4

Average lexical densities at this level were 43.4 ($SD=3.9$) and 47.9 ($SD=3.0$) for Luther and Hfa respectively. The differences were significant ($t(4)=-8.36, p < .05$), using a two-tailed t-test.
The final statistical comparison occurs at two levels: between each of the five books, and between the complete texts:

Table 3.5

<table>
<thead>
<tr>
<th></th>
<th>Lexical Density</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Luther</td>
<td>Hfa</td>
</tr>
<tr>
<td>Matthew</td>
<td>50.5</td>
<td>53.1</td>
</tr>
<tr>
<td>Mark</td>
<td>46.8</td>
<td>53.7</td>
</tr>
<tr>
<td>Luke</td>
<td>48.9</td>
<td>54.0</td>
</tr>
<tr>
<td>John</td>
<td>49.5</td>
<td>52.5</td>
</tr>
<tr>
<td>Acts</td>
<td>50.2</td>
<td>54.0</td>
</tr>
<tr>
<td>Overall</td>
<td>49.4</td>
<td>53.5</td>
</tr>
</tbody>
</table>

As was seen in the comparison of the chapters, Luther is once again consistently less dense than Hfa, as is shown in the figure below:

Figure 3.5

Figure 3.5: Lexical Density per book and overall
In this case, the sample sizes are the entire texts, with lengths of 101,288 words for the
Luther text and 102,602 words for Hfa, for a difference in total length of 1,314 words.
The difference in the number of lexical words is more considerable, with Hfa having
4,871 more lexical words than Luther. This value constitutes 8.89% of the total number
of lexical words in Hfa. Average lexical densities over the books are 49.2 for Luther and
53.4 for lexical density. This can then be compared to the calculated lexical density for
each text which is 49.4 ($SD=1.3$) for Luther and 53.5 ($SD=0.6$) for Hfa. The differences
between the two Bible versions are once again significant ($t(5)=-6.61$, $p < .05$), using a
two-tailed t-test.

The removal of grammatical verbs from the calculation does nothing to alter the
relative differences in lexical density at this level either, as can be seen in the next table:

Table 3.6

<table>
<thead>
<tr>
<th></th>
<th>Lexical Density</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Luther</td>
</tr>
<tr>
<td>Matthew</td>
<td>44.0</td>
</tr>
<tr>
<td>Mark</td>
<td>41.4</td>
</tr>
<tr>
<td>Luke</td>
<td>42.3</td>
</tr>
<tr>
<td>John</td>
<td>41.9</td>
</tr>
<tr>
<td>Acts</td>
<td>44.8</td>
</tr>
<tr>
<td>Overall</td>
<td>43.0</td>
</tr>
</tbody>
</table>

As before, at the most lengthy level of analysis, Luther continues to be less dense, still
containing a relatively lower number of lexical words:

Figure 3.6
At this level Luther has an average lexical density of 42.9 (SD=1.3) and Hfa has 46.6 (SD=1.2). The differences are significant ($t(5)=-6.83$, $p <.05$), using a two-tailed t-test. In looking at all lexical density stats, both with auxiliaries and without, the t-test again revealed significant results ($t(41)=-8.71$, $p <.05$).

It is notable as well that, while Hfa shows consistently higher lexical density than Luther, the average difference in lexical densities decreases as sample length increases. At the verse level, the average difference in lexical densities is 8.1, while at the chapter level it decreases to 5.0, followed by 4.3 at the book and overall level. With modal and auxiliary verbs removed the differences are similar, albeit less pronounced, with a difference of 7.8 at the verse level, 4.5 at the chapter level, and 3.7 at the level of book and complete corpus.
3.3.2 Lexical Word Types

In order to further investigate the differences in lexical density, the concentration of the two main types of lexical words – nouns and verbs – is examined in this section. Nouns and verbs make up 80.7 percent of the lexical words in Luther and 79.2 percent of the lexical words in Hfa. While the total percentages are similar, the breakdown of the word types at various levels is rather different, as will be seen in the sections below. The first type to be compared is the concentration of nouns, followed by the concentration of verbs.

3.3.2.1 Concentration of Nouns

The concentration of nouns, as can be expected, varies widely from verse to verse, from no nouns being present at all in the Luther version of Luke 18:26 to nouns comprising more than half of all lexical words in Hfa’s John 13:5.

Table 3.7
As can be seen in table 3.4 and figure 3.4 below, in eight out of ten examples, Luther has a lower percentage of nouns, even to the point of having no nouns at all in one passage. The differences are statistically significant with a two-tailed t-test ($t(9)=-2.89, p < .05$).

**Figure 3.7**

<table>
<thead>
<tr>
<th></th>
<th>Percent Nouns</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Luther</td>
<td></td>
<td>Hfa</td>
</tr>
<tr>
<td>Matthew 8:13</td>
<td>36.4</td>
<td></td>
<td>37.5</td>
</tr>
<tr>
<td>Matthew 15:9</td>
<td>33.3</td>
<td></td>
<td>50.0</td>
</tr>
<tr>
<td>Mark 7:11</td>
<td>40.0</td>
<td></td>
<td>31.3</td>
</tr>
<tr>
<td>Mark 14:21</td>
<td>22.2</td>
<td></td>
<td>20.0</td>
</tr>
<tr>
<td>Luke 4:6</td>
<td>20.0</td>
<td></td>
<td>41.7</td>
</tr>
<tr>
<td>Luke 18:26</td>
<td>0.0</td>
<td></td>
<td>12.5</td>
</tr>
<tr>
<td>John 6:13</td>
<td>25.0</td>
<td></td>
<td>30.8</td>
</tr>
<tr>
<td>John 13:5</td>
<td>41.7</td>
<td></td>
<td>55.6</td>
</tr>
<tr>
<td>Acts 11:23</td>
<td>28.6</td>
<td></td>
<td>43.8</td>
</tr>
<tr>
<td>Acts 16:39</td>
<td>16.7</td>
<td></td>
<td>45.5</td>
</tr>
</tbody>
</table>

Table 3.7: Percent Nouns per Verse
While the number of nouns does not decrease when modal and auxiliary verbs are removed from the calculations, their removal does increase the relative concentration of nouns within the group of lexical words. This increase at the verse level is clear in the following table:

Table 3.8

<table>
<thead>
<tr>
<th>Verse</th>
<th>Luther</th>
<th>Hfa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matthew 8:13</td>
<td>44.4</td>
<td>50.0</td>
</tr>
<tr>
<td>Matt 15:9</td>
<td>40.0</td>
<td>57.1</td>
</tr>
<tr>
<td>Mark 7:11</td>
<td>50.0</td>
<td>41.7</td>
</tr>
<tr>
<td>Mark 14:21</td>
<td>26.7</td>
<td>27.3</td>
</tr>
<tr>
<td>Luke 4:6</td>
<td>28.6</td>
<td>50.0</td>
</tr>
<tr>
<td>Luke 18:26</td>
<td>0.0</td>
<td>16.7</td>
</tr>
<tr>
<td>John 6:13</td>
<td>27.3</td>
<td>33.3</td>
</tr>
<tr>
<td>John 13:5</td>
<td>45.5</td>
<td>55.6</td>
</tr>
<tr>
<td>Acts 11:23</td>
<td>33.3</td>
<td>46.7</td>
</tr>
<tr>
<td>Acts 16:39</td>
<td>16.7</td>
<td>45.5</td>
</tr>
</tbody>
</table>

The differences between the two Bible texts are again statistically significant with a two-tailed t-test ($t(9)=-3.28, p <.05$). The following chart shows that the removal of modals and auxiliaries from the group of lexical words changes only one of the relative concentrations and comparisons:

Figure 3.8
Without modals and auxiliaries, the example from Mark 14:21 now has a lower concentration of nouns in Luther than the version from Hfa. Otherwise, the comparisons of higher and lower noun concentration remain the same.

As the length of the samples increases, we see a similar pattern. At the chapter level, all but one of the chapters in Luther contain fewer nouns. The differences here, however, are not statistically significant ($t(4) = -1.63, p < .05$).

Table 3.9

<table>
<thead>
<tr>
<th></th>
<th>Percent Nouns</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Luther</td>
<td>Hfa</td>
</tr>
<tr>
<td>Matthew 4</td>
<td>45.6</td>
<td>48.6</td>
</tr>
<tr>
<td>Mark 6</td>
<td>36.2</td>
<td>40.1</td>
</tr>
<tr>
<td>Luke 15</td>
<td>30.8</td>
<td>30.9</td>
</tr>
<tr>
<td>John 1</td>
<td>41.8</td>
<td>40.7</td>
</tr>
<tr>
<td>Acts 11</td>
<td>41.2</td>
<td>43.0</td>
</tr>
</tbody>
</table>
While the pattern does continue, the differences are less pronounced, as can be clearly seen in figure 3.9 below. Luke 15 is almost even, and even where Luther exceeds Hfa in John 1, it is only by 1.1 percent.

Figure 3.9

![Bar chart showing percent nouns per chapter for Matthew 4, Mark 6, Luke 15, John 1, and Acts 11 between Luther and Hfa]

At the chapter level we again see a slight increase in some areas in the relative concentration of nouns in Luther with the removal of modals and auxiliaries:

Table 3.10

<table>
<thead>
<tr>
<th></th>
<th>Percent Nouns</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Luther</td>
</tr>
<tr>
<td>Matthew 4</td>
<td>48.4</td>
</tr>
<tr>
<td>Mark 6</td>
<td>40.4</td>
</tr>
<tr>
<td>Luke 15</td>
<td>36.8</td>
</tr>
<tr>
<td>John 1</td>
<td>49.6</td>
</tr>
<tr>
<td>Acts 11</td>
<td>45.4</td>
</tr>
</tbody>
</table>
Relative to Hfa, we see that where the levels were about equal in the chapter from Luke and just slightly higher in the example from John, the levels are now clearly higher in Luke for Luther and still higher again in John in the Luther text. Once again, the differences are not statistically significant ($t(4)= -1.07, p <.05$). This can be seen more clearly in the chart below:

**Figure 3.10**

![Figure 3.10: Percent Nouns per Chapter without Modals and Auxiliaries](image)

At the level of book and the overall corpus, we see the pattern of Luther containing fewer nouns continue, except in the case of the book of John, as can be seen below in table 3.11:

**Table 3.11**
Table 3.11: Percent Nouns per Book and Entire Corpus

<table>
<thead>
<tr>
<th></th>
<th>Percent Nouns</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Luther</td>
<td>Hfa</td>
</tr>
<tr>
<td>Matthew</td>
<td>36.4</td>
<td>38.8</td>
</tr>
<tr>
<td>Mark</td>
<td>35.5</td>
<td>38.5</td>
</tr>
<tr>
<td>Luke</td>
<td>35.8</td>
<td>37.9</td>
</tr>
<tr>
<td>John</td>
<td>33.9</td>
<td>33.5</td>
</tr>
<tr>
<td>Acts</td>
<td>40.1</td>
<td>42.3</td>
</tr>
<tr>
<td>Overall</td>
<td>36.5</td>
<td>38.4</td>
</tr>
</tbody>
</table>

As the length of sample increases, the differences between the two versions decrease, as can be seen in the figure below. Still, it is clear that Luther typically contains a lower concentration of nouns. The differences here are found to be statistically significant ($t(5) = -3.83, p < .05$).

Figure 3.11

Figure 3.11: Percent Nouns per Book and Entire Corpus
The differences between the two texts do seem to become more marked with the removal of auxiliaries and modals at the level of book and the complete corpus, as illustrated in table 3.12:

Table 3.12

<table>
<thead>
<tr>
<th></th>
<th>Percent Nouns</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Luther</td>
<td>Hfa</td>
</tr>
<tr>
<td>Matthew</td>
<td>41.8</td>
<td>44.5</td>
</tr>
<tr>
<td>Mark</td>
<td>40.1</td>
<td>43.9</td>
</tr>
<tr>
<td>Luke</td>
<td>41.4</td>
<td>43.6</td>
</tr>
<tr>
<td>John</td>
<td>40.1</td>
<td>39.6</td>
</tr>
<tr>
<td>Acts</td>
<td>44.9</td>
<td>47.5</td>
</tr>
<tr>
<td>Overall</td>
<td>41.9</td>
<td>44.0</td>
</tr>
</tbody>
</table>

Once again, other than in the book of John, Luther is consistently lower in concentration of nouns. This is confirmed by a two-tailed t-test, which indicates a significant difference ($t(5)=-3.72$, $p <.05$). The differences are now greater, ranging from 0.5% to 3.8%, whereas when modals and auxiliaries were included, the differences at this level only went as high as 2.4%. This can be more clearly visualized with figure 3.12:

Figure 3.12
3.3.2.2 Concentration of Verbs

In looking at the concentration of verbs at the different levels of analysis, in most cases we see a pattern opposite of that found with noun concentration. That is, Luther typically contains the higher concentration of verbs. At the level of the verse, we can see below that in the selected verses, eight of ten follow this pattern (see table 3.13):

Table 3.13
In some cases, as can be seen in figure 3.13, the difference is great, as in the example from Acts 16:39. Statistically, we see a significant difference, once again using the two-tailed t-test ($t(9)= 2.81, p < .05$).

Figure 3.13
The removal of modal and auxiliary verbs drastically changes the concentrations for all levels of analysis. The following table shows this difference at the verse level:

Table 3.14

<table>
<thead>
<tr>
<th></th>
<th>Percent Verbs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Luther</td>
</tr>
<tr>
<td>Matthew 8:13</td>
<td>44.4</td>
</tr>
<tr>
<td>Matt 15:9</td>
<td>40.0</td>
</tr>
<tr>
<td>Mark 7:11</td>
<td>50.0</td>
</tr>
<tr>
<td>Mark 14:21</td>
<td>33.3</td>
</tr>
<tr>
<td>Luke 4:6</td>
<td>57.1</td>
</tr>
<tr>
<td>Luke 18:26</td>
<td>40.0</td>
</tr>
<tr>
<td>John 6:13</td>
<td>36.4</td>
</tr>
<tr>
<td>John 13:5</td>
<td>45.5</td>
</tr>
<tr>
<td>Acts 11:23</td>
<td>41.7</td>
</tr>
<tr>
<td>Acts 16:39</td>
<td>83.3</td>
</tr>
</tbody>
</table>

Concentrations of verbs are clearly lower across the board. In most cases, however, Luther continues to have a higher concentration, as evidenced by the comparisons in the following chart:

Figure 3.14
Differences range from 1.0% to as high as 25.7% at this level. These differences are significant as well ($t(9) = 2.74$, $p < .05$).

At the level of chapter, the analysis shows that in all cases, Luther has the higher concentration of verbs, with differences ranging from 2.1% to 8.1%.

Table 3.15

<table>
<thead>
<tr>
<th></th>
<th>Percent Verbs</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Luther</td>
<td>Hfa</td>
<td></td>
</tr>
<tr>
<td>Matthew 4</td>
<td>36.3</td>
<td>34.2</td>
<td></td>
</tr>
<tr>
<td>Mark 6</td>
<td>45.7</td>
<td>38.4</td>
<td></td>
</tr>
<tr>
<td>Luke 15</td>
<td>48.8</td>
<td>40.7</td>
<td></td>
</tr>
<tr>
<td>John 1</td>
<td>46.8</td>
<td>40.7</td>
<td></td>
</tr>
<tr>
<td>Acts 11</td>
<td>38.0</td>
<td>34.1</td>
<td></td>
</tr>
</tbody>
</table>
However, the figure below shows that the differences between the samples are much less extreme at this level than at the level of verse. They do, however, remain significant ($t(4)= 4.99, p < .05$).

Figure 3.15

![Figure 3.15: Percent Verbs per Chapter](image)

Without modals and auxiliaries, Luther continues to have the higher concentration of verbs. This time, differences in concentration range from 2.1%-9.2%. This difference is significant ($t(4)= 5.09$). The consistency of this pattern can be seen below in table 3.16 and figure 3.16:

Table 3.16
At the level of entire books, this pattern is born out once again. This time the differences range from 2.1% to 5.3%, and are significant ($t(5) = 7.49$, $p < .05$).

Table 3.17
The pattern that we see below in figure 3.17 is consistent throughout the different levels of analysis. In all cases and at all levels, Luther has a higher concentration of verbs.

Figure 3.17

<table>
<thead>
<tr>
<th>Book</th>
<th>Luther</th>
<th>Hfa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matthew</td>
<td>43.3</td>
<td>41.1</td>
</tr>
<tr>
<td>Mark</td>
<td>45.7</td>
<td>40.4</td>
</tr>
<tr>
<td>Luke</td>
<td>45.1</td>
<td>41.2</td>
</tr>
<tr>
<td>John</td>
<td>46.9</td>
<td>44.6</td>
</tr>
<tr>
<td>Acts</td>
<td>40.7</td>
<td>37.4</td>
</tr>
<tr>
<td>Overall</td>
<td>44.1</td>
<td>40.8</td>
</tr>
</tbody>
</table>

Table 3.17: Percent Verbs per Book and Entire Corpus
After the removal of modals and auxiliaries at this level, the pattern remains the same, with Luther still having a higher concentration of verbs than Hfa, as evidenced by table 3.18 and figure 3.18:

Table 3.18

<table>
<thead>
<tr>
<th></th>
<th>Percent Verbs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Luther</td>
</tr>
<tr>
<td>Matthew</td>
<td>34.9</td>
</tr>
<tr>
<td>Mark</td>
<td>38.6</td>
</tr>
<tr>
<td>Luke</td>
<td>36.5</td>
</tr>
<tr>
<td>John</td>
<td>37.3</td>
</tr>
<tr>
<td>Acts</td>
<td>33.6</td>
</tr>
<tr>
<td>Overall</td>
<td>35.9</td>
</tr>
</tbody>
</table>

Figure 3.18

Figure 3.18: Percent Verbs per Book and Entire Corpus without Modals and Auxiliaries
The differences at this level, even without modals and auxiliaries, continue to be significant ($t(5)= 6.90, p < .05$). With this being the case, a key point of the discussion later in this chapter will be to examine whether or not there is a correlation between concentrations of nouns and verbs and lexical density overall.

3.4 Discussion

This section is divided into three parts: an overview of the lexical patterns found in the texts, an analysis of the statistics pertaining to noun and verb concentrations, and finally a discussion of how these measures fit together and pertain to readability.

3.4.1 Lexical Density Patterns

In looking at patterns of lexical density in the samples, it is helpful to first have some examples from individual verses. The first example comes from the book of Matthew:

**Matthew 15:9 (Luther):** vergeblich dienen sie mir, weil sie **lehren** solche **Lehren**, die nichts als **Menschengebote sind**.

**Matthew 15:9 (Hfa):** Ihr **Gottesdienst ist wertlos**, weil sie ihre menschlichen **Gesetze** als **Gebote Gottes ausgeben**.

The above examples from Matthew show a pattern that is fairly common in this comparison. As seen above in the data analysis section, this example is part of a pattern where the majority of compared verses show lower lexical density in Luther than in Hfa.
In this case, the verses are of similar length. In the next example, however, the verse from Hfa shows both higher lexical density and considerably greater length:

**Mark 7:11 (Luther):** Ihr aber **lehrt:** Wenn einer zu **Vater** oder **Mutter** sagt: **Korban** - das **heißt:** **Opfergabe** soll sein, was dir von mir **zusteht**

**Mark 7:11 (Hfa):** Ihr aber **behauptet:** Wenn jemand seinen **hilfsbedürftigen Eltern erklärt:** 'Ich kann euch nicht **helfen,** weil ich mein **Vermögen** dem **Tempel vermacht** habe', dann hat er nicht gegen **Gottes Gebot** **verstoßen.**

The above example from Hfa has added a great deal more explanation to the text, including many more lexical words. In the Luther example, there are a total of 21 words, eight of which are lexical words, consisting of four nouns and four verbs. In contrast, the Hfa text has 30 words, 13 of which are lexical words, this time containing five nouns, an adjective, five verbs and two adverbs.

The final example from the individual verses examined is the only one in which Luther is denser than Hfa:

**John 13:5 (Luther):** Darnach **goss** er **Wasser** in ein **Becken,** hob an, den **Jüngern** die **Füße zu waschen,** und **trocknete** sie mit dem **Schurz,** damit er **umgürtet** war.

**John 13:5 (Hfa):** Er **goss** **Wasser** in eine **Schüssel** und **begann,** seinen **Jüngern** die **Füße zu waschen** und mit dem **Tuch abzutrocknen.**

This time, the example from Luther has 25 total words, with five nouns, five verbs and one adverb. The Hfa verse is shorter at 19 words, with five nouns and four verbs. This is, in fact, the only time in all of the data examined where Luther is denser than Hfa.

Due to the nature of the tag set, lexical density was calculated twice, to make sure that the inclusion of auxiliaries and modals in the calculation of lexical density did not lead to a false pattern. Historical developments in German could lead one to believe that
this may have been the case. According to Sommerfeldt (1988), as German has developed from earlier Germanic forms, the number of synthetic tenses – tenses that rely on inflection – has been reduced to two: present and simple past (preterit) (p. 209). In contrast, there are three analytic tenses: present perfect, past perfect and future (ibid.). One could, thus, make an assumption that the higher lexical density calculated for Hfa was due to the inclusion of modals and auxiliaries used in the analytic tenses, since German has more of these tenses. The recalculation of lexical density and noun and verb concentration without grammatical verbs, however, proves that this is not the case. Even without such verbs, Luther continues to show lower lexical density.

As we can see from the data above, at all levels Luther shows lower overall lexical density. In looking at the hypotheses for this thesis, it appears from these statistics that the Luther text was more readable for its contemporaries than Hfa is for readers of the late 20th and early 21st centuries. The statistics of lexical density, however, while they focus on lexical words in general, do not give a picture of the relative concentrations of different parts of speech. As seen in the data and discussion above, there are differences in the concentrations of nouns and verbs in the text as well. The next section will discuss this issue further.

3.4.2 Differences in Types of Lexical Words

There are interesting patterns to be seen in the concentration of nouns and verbs in the texts. From the standpoint of nouns, it is not totally consistent whether one version has a higher concentration than the other. At the verse level, most of the verses compared
show a lower concentration of nouns in the Luther corpus. Both verses from Mark, however, have a higher concentration of nouns compared to Hfa. In contrast, the second example from Luke – verse 18:26 – has no nouns at all. Similar inconsistencies continue at the chapter level. Luke 15 is virtually equal in both Bible versions, and John 1 has a higher concentration of nouns in Luther. At book level, Luther has a lower concentration of nouns than Hfa for all books except John.

At the level of the verse, concentration of verbs also appears to be inconsistent, this time with Luther generally having the higher concentration of verbs than Hfa except for the examples of Mark 14:21 and Luke 18:26.

An important question for this section is whether or not there is a significant correlation between the percentages of nouns and verbs and lexical density. In examining all of the data together, a two-tailed correlation test showed no significant correlation between lexical density and the percentage of nouns in the set of lexical words (N=42, correlation coefficient = -.271, α = .05, critical value = .304 or -.304). There is, however, a significant negative correlation between lexical density and the percentage of verbs (N=42, correlation coefficient = -.377, α = .05, critical value = .304 or -.304). The trend we see here is striking – as lexical density increases, the concentration of verbs in the passage decreases. More importantly, we can assume then that increasing the concentration of verbs in a text lowers its lexical density. This becomes even more dramatic when modal and auxiliary verbs are removed from the calculation. In this case, the results are highly significant, with a correlation coefficient of -0.661 (N=42 α = .05, critical value = .304 or -.304).
This is further evidence that the Luther text was translated in a more readable manner than Hfa. While little can be said for the level of nouns in either text, it is clear from the data that Luther consistently had a higher concentration of verbs, making it less dense.

3.5 Conclusion

The lexical density and lexical word concentration data seen in this chapter provide us with two important observations. First and foremost, the Luther text has consistently lower lexical density than Hfa. This is coupled with the second observation that lower lexical density is paired with a higher concentration of verbs within the set of lexical words found in the text. We can conclude from this that, while both texts have the goal of being readable for their contemporaries, Luther is the more successful text. It is also important to understand what sort of shift in language makes Hfa more lexically dense. This comes, according to the statistics, from a decrease in the use of verbs in the sentences. Verb frequency is also important for clausal structure, a topic that will be discussed in the next chapter. While clauses in German may contain an unlimited number of lexical words such as nouns, adjectives and adverbs, the German clause is limited in the number of verbs it may contain, typically being limited to one finite verb and one to two other verb forms, such as participles and infinitives. While not looking specifically at the different types of verbs, one can make the claim that a higher concentration of verbs also indicates a greater number of clauses in the text. This will be further corroborated in the next section, which examines markers of clausal complexity.
Chapter 4
Clausal Complexity

4.1 Introduction

This chapter looks at how clauses are bound together, specifically measuring the frequency of different conjunction types. While lexical density in this project is not measured with reference to clauses or clausal structure, the clause is still an important feature of the text, as outlined in chapter two. The research questions for this chapter focus on two areas: 1) the frequency of different types of conjunctions and 2) the relationship of that frequency to lexical density.

As discussed in chapter 2, two types of clausal relations are important for this study – paratactic and hypotactic. Paratactic clausal relations involve “the linking of elements of equal status” (Halliday and Matthiessen 2004, p. 384). This often involves the use of coordinating conjunctions, such as und “and”, and oder “or”. Hypotaxis, on the other hand, indicates “the binding of elements of unequal status” (p. 387). In German, such clauses are typically started with subordinating conjunctions, such as dass “that”, and weil “because”. Examples of these clausal relations in German can be seen below:

Ich bin müde, und ich bin krank. (I am tired, and I am sick.)
Ich bin müde, weil ich krank bin. (I am tired, because I am sick.)

Such clausal relations relate to readability in that greater use of conjunctions indicates that a text has greater grammatical intricacy. Grammatical intricacy, in turn, is
a hallmark of spoken language, which is less lexically dense than written language. Therefore, one translated text that is more grammatically intricate than another translation of the same text, should, in turn, be less dense, and thus more readable. The measurement to be used for this is the frequency of different types of conjunctions, as will be discussed below.

4.2 Methodology

This chapter utilizes the same corpus and POS tagging system seen in chapter 3, this time focusing on different varieties of conjunctions. The STTS (Schiller, Teufel et al. 1995) contains several different types of conjunction tags. For the purpose of this chapter, two types were used: KON, which are coordinating conjunctions, and KOUS, which are subordinating conjunctions. The analysis took place in two ways with the coordinating conjunctions, once with the conjunction *und* “and” included and once without. This is due to the overwhelming frequency of this conjunction in comparison to the other types, and the fact that *und* “and” can often be used as part of a list, rather than as a conjunction. As with the modal and auxiliary verbs in the previous chapter, this was done to ensure that their inclusion or exclusion did not affect the relative results of the study. In order to support this, several correlations will be shown, both including *und* “and” and excluding it. The correlations will be run with both measures of lexical density seen in the previous chapter.
4.3 Data Analysis

The analysis of the data for this chapter proceeds in four sections: 1) frequency of different types of conjunctions; 2) comparative frequencies of coordinating conjunctions; 3) comparative frequencies of subordinating conjunctions; 4) correlation analysis of conjunction concentration with lexical density.

4.3.1 Frequency of Different Conjunction Types

The data analysis begins by looking at the different types of conjunctions found in the text and their frequencies, in order to ascertain if any conjunctions stand out. Looking first at coordinating conjunctions, table 4.1 shows the coordinating conjunctions found in Luther, including the word form, its corpus tag and the lemma to which it belongs:

Table 4.1
### Table 4.1: Coordinating Conjunctions in Luther

<table>
<thead>
<tr>
<th>frequency</th>
<th>token</th>
<th>tag</th>
<th>lemma</th>
</tr>
</thead>
<tbody>
<tr>
<td>5569</td>
<td>und</td>
<td>KON</td>
<td>und</td>
</tr>
<tr>
<td>1531</td>
<td>Und</td>
<td>KON</td>
<td>und</td>
</tr>
<tr>
<td>390</td>
<td>denn</td>
<td>KON</td>
<td>denn</td>
</tr>
<tr>
<td>299</td>
<td>aber</td>
<td>KON</td>
<td>aber</td>
</tr>
<tr>
<td>205</td>
<td>Denn</td>
<td>KON</td>
<td>denn</td>
</tr>
<tr>
<td>174</td>
<td>sondern</td>
<td>KON</td>
<td>sondern</td>
</tr>
<tr>
<td>143</td>
<td>Aber</td>
<td>KON</td>
<td>aber</td>
</tr>
<tr>
<td>132</td>
<td>oder</td>
<td>KON</td>
<td>oder</td>
</tr>
<tr>
<td>86</td>
<td>bis</td>
<td>KON</td>
<td>bis</td>
</tr>
<tr>
<td>27</td>
<td>noch</td>
<td>KON</td>
<td>noch</td>
</tr>
<tr>
<td>21</td>
<td>weder</td>
<td>KON</td>
<td>weder</td>
</tr>
<tr>
<td>19</td>
<td>Doch</td>
<td>KON</td>
<td>doch</td>
</tr>
<tr>
<td>16</td>
<td>Oder</td>
<td>KON</td>
<td>oder</td>
</tr>
<tr>
<td>8</td>
<td>doch</td>
<td>KON</td>
<td>doch</td>
</tr>
<tr>
<td>6</td>
<td>als</td>
<td>KON</td>
<td>als</td>
</tr>
<tr>
<td>4</td>
<td>wie</td>
<td>KON</td>
<td>wie</td>
</tr>
<tr>
<td>3</td>
<td>entweder</td>
<td>KON</td>
<td>entweder</td>
</tr>
<tr>
<td>3</td>
<td>desto</td>
<td>KON</td>
<td>desto</td>
</tr>
<tr>
<td>3</td>
<td>Indem</td>
<td>KON</td>
<td>indem</td>
</tr>
<tr>
<td>1</td>
<td>sowohl</td>
<td>KON</td>
<td>sowohl</td>
</tr>
<tr>
<td>1</td>
<td>Weil</td>
<td>KON</td>
<td>weil</td>
</tr>
</tbody>
</table>

The first thing of note is the overwhelming frequency of the conjunction *und* “and”, which was noted above in the methodology section. It appears either capitalized or lowercase over 7,000 times in the Luther text. The next highest conjunction *denn* “because” only appears 390 times. There are some members of the above chart that are not canonically viewed as coordinating conjunctions. *Doch* is counted in this category due to its function according to the tag set as a sentence beginning coordinating
conjunction (Schiller, et al. 1995, p. 59). The other non-canon conjunctions are part of the list due to idiosyncrasies in the programming of the tagger.

The frequency of coordinating conjunctions in Hfa appears as follows:

Table 4.2

<table>
<thead>
<tr>
<th>frequency</th>
<th>token</th>
<th>tag</th>
<th>lemma</th>
</tr>
</thead>
<tbody>
<tr>
<td>3108</td>
<td>und</td>
<td>KON</td>
<td>und</td>
</tr>
<tr>
<td>372</td>
<td>Aber</td>
<td>KON</td>
<td>aber</td>
</tr>
<tr>
<td>225</td>
<td>Und</td>
<td>KON</td>
<td>und</td>
</tr>
<tr>
<td>151</td>
<td>Denn</td>
<td>KON</td>
<td>denn</td>
</tr>
<tr>
<td>146</td>
<td>aber</td>
<td>KON</td>
<td>aber</td>
</tr>
<tr>
<td>145</td>
<td>denn</td>
<td>KON</td>
<td>denn</td>
</tr>
<tr>
<td>129</td>
<td>oder</td>
<td>KON</td>
<td>oder</td>
</tr>
<tr>
<td>124</td>
<td>Doch</td>
<td>KON</td>
<td>doch</td>
</tr>
<tr>
<td>81</td>
<td>sondern</td>
<td>KON</td>
<td>sondern</td>
</tr>
<tr>
<td>60</td>
<td>bis</td>
<td>KON</td>
<td>bis</td>
</tr>
<tr>
<td>42</td>
<td>Weil</td>
<td>KON</td>
<td>weil</td>
</tr>
<tr>
<td>31</td>
<td>doch</td>
<td>KON</td>
<td>doch</td>
</tr>
<tr>
<td>21</td>
<td>weder</td>
<td>KON</td>
<td>weder</td>
</tr>
<tr>
<td>19</td>
<td>Oder</td>
<td>KON</td>
<td>oder</td>
</tr>
<tr>
<td>12</td>
<td>Noch</td>
<td>KON</td>
<td>noch</td>
</tr>
<tr>
<td>9</td>
<td>wie</td>
<td>KON</td>
<td>wie</td>
</tr>
<tr>
<td>9</td>
<td>noch</td>
<td>KON</td>
<td>noch</td>
</tr>
<tr>
<td>6</td>
<td>sowie</td>
<td>KON</td>
<td>sowie</td>
</tr>
<tr>
<td>6</td>
<td>als</td>
<td>KON</td>
<td>als</td>
</tr>
<tr>
<td>2</td>
<td>sowohl</td>
<td>KON</td>
<td>sowohl</td>
</tr>
<tr>
<td>1</td>
<td>entweder</td>
<td>KON</td>
<td>entweder</td>
</tr>
<tr>
<td>1</td>
<td>Weder</td>
<td>KON</td>
<td>weder</td>
</tr>
<tr>
<td>1</td>
<td>Entweder</td>
<td>KON</td>
<td>entweder</td>
</tr>
</tbody>
</table>

Once again, lower cased und “and” is the most frequent by far, with 3,108 occurrences, followed distantly by Aber “but” in uppercase at 372 occurrences. Unlike Luther,
capitalized Und “and” only appears in third place with 225 occurrences in Hfa, compared to the 1531 occurrences in Luther.

Subordinating conjunctions in both texts show greater variety, as can be seen in the following table from Luther:

Table 4.3

<table>
<thead>
<tr>
<th>frequency</th>
<th>token</th>
<th>tag</th>
<th>lemma</th>
</tr>
</thead>
<tbody>
<tr>
<td>1187</td>
<td>daß</td>
<td>KOUS</td>
<td>daß</td>
</tr>
<tr>
<td>150</td>
<td>da</td>
<td>KOUS</td>
<td>da</td>
</tr>
<tr>
<td>115</td>
<td>wenn</td>
<td>KOUS</td>
<td>wenn</td>
</tr>
<tr>
<td>113</td>
<td>Da</td>
<td>KOUS</td>
<td>da</td>
</tr>
<tr>
<td>76</td>
<td>wie</td>
<td>KOUS</td>
<td>wie</td>
</tr>
<tr>
<td>68</td>
<td>Wenn</td>
<td>KOUS</td>
<td>wenn</td>
</tr>
<tr>
<td>37</td>
<td>als</td>
<td>KOUS</td>
<td>als</td>
</tr>
<tr>
<td>35</td>
<td>ob</td>
<td>KOUS</td>
<td>ob</td>
</tr>
<tr>
<td>19</td>
<td>ehe</td>
<td>KOUS</td>
<td>ehe</td>
</tr>
<tr>
<td>13</td>
<td>nachdem</td>
<td>KOUS</td>
<td>nachdem</td>
</tr>
<tr>
<td>13</td>
<td>dieweil</td>
<td>KOUS</td>
<td>dieweil</td>
</tr>
<tr>
<td>8</td>
<td>weil</td>
<td>KOUS</td>
<td>weil</td>
</tr>
<tr>
<td>8</td>
<td>indem</td>
<td>KOUS</td>
<td>indem</td>
</tr>
<tr>
<td>8</td>
<td>Als</td>
<td>KOUS</td>
<td>als</td>
</tr>
<tr>
<td>7</td>
<td>Wie</td>
<td>KOUS</td>
<td>wie</td>
</tr>
<tr>
<td>6</td>
<td>Daß</td>
<td>KOUS</td>
<td>daß</td>
</tr>
<tr>
<td>4</td>
<td>wiewohl</td>
<td>KOUS</td>
<td>wiewohl</td>
</tr>
<tr>
<td>4</td>
<td>solange</td>
<td>KOUS</td>
<td>solange</td>
</tr>
<tr>
<td>3</td>
<td>Nachdem</td>
<td>KOUS</td>
<td>nachdem</td>
</tr>
<tr>
<td>1</td>
<td>sobald</td>
<td>KOUS</td>
<td>sobald</td>
</tr>
<tr>
<td>1</td>
<td>nun</td>
<td>KOUS</td>
<td>nun</td>
</tr>
<tr>
<td>1</td>
<td>damit</td>
<td>KOUS</td>
<td>damit</td>
</tr>
<tr>
<td>1</td>
<td>bis</td>
<td>KOUS</td>
<td>bis</td>
</tr>
<tr>
<td>1</td>
<td>Solange</td>
<td>KOUS</td>
<td>solange</td>
</tr>
<tr>
<td>1</td>
<td>Ob</td>
<td>KOUS</td>
<td>ob</td>
</tr>
<tr>
<td>1</td>
<td>Nun</td>
<td>KOUS</td>
<td>nun</td>
</tr>
<tr>
<td>1</td>
<td>Je</td>
<td>KOUS</td>
<td>Je</td>
</tr>
</tbody>
</table>
Luther shows a preference for one conjunction in particular – *dass* “that” – with 1,187 occurrences. The next most frequent conjunction is *da* “because”, with 150 occurrences.

Similarly, Hfa shows greater variety of subordinating conjunctions:

Table 4.4
Table 4.4: Subordinating Conjunctions in Hfa

<table>
<thead>
<tr>
<th>frequency</th>
<th>token</th>
<th>tag</th>
<th>lemma</th>
</tr>
</thead>
<tbody>
<tr>
<td>605</td>
<td>daß</td>
<td>KOUS</td>
<td>daß</td>
</tr>
<tr>
<td>252</td>
<td>Wenn</td>
<td>KOUS</td>
<td>wenn</td>
</tr>
<tr>
<td>181</td>
<td>wenn</td>
<td>KOUS</td>
<td>wenn</td>
</tr>
<tr>
<td>154</td>
<td>weil</td>
<td>KOUS</td>
<td>weil</td>
</tr>
<tr>
<td>105</td>
<td>wie</td>
<td>KOUS</td>
<td>wie</td>
</tr>
<tr>
<td>53</td>
<td>Nachdem</td>
<td>KOUS</td>
<td>nachdem</td>
</tr>
<tr>
<td>25</td>
<td>obwohl</td>
<td>KOUS</td>
<td>obwohl</td>
</tr>
<tr>
<td>25</td>
<td>Als</td>
<td>KOUS</td>
<td>als</td>
</tr>
<tr>
<td>24</td>
<td>damit</td>
<td>KOUS</td>
<td>damit</td>
</tr>
<tr>
<td>22</td>
<td>nachdem</td>
<td>KOUS</td>
<td>nachdem</td>
</tr>
<tr>
<td>18</td>
<td>als</td>
<td>KOUS</td>
<td>als</td>
</tr>
<tr>
<td>16</td>
<td>ob</td>
<td>KOUS</td>
<td>ob</td>
</tr>
<tr>
<td>16</td>
<td>bevor</td>
<td>KOUS</td>
<td>bevor</td>
</tr>
<tr>
<td>12</td>
<td>ehe</td>
<td>KOUS</td>
<td>ehe</td>
</tr>
<tr>
<td>10</td>
<td>Während</td>
<td>KOUS</td>
<td>während</td>
</tr>
<tr>
<td>10</td>
<td>Wie</td>
<td>KOUS</td>
<td>wie</td>
</tr>
<tr>
<td>9</td>
<td>solange</td>
<td>KOUS</td>
<td>solange</td>
</tr>
<tr>
<td>8</td>
<td>während</td>
<td>KOUS</td>
<td>während</td>
</tr>
<tr>
<td>5</td>
<td>Bevor</td>
<td>KOUS</td>
<td>bevor</td>
</tr>
<tr>
<td>4</td>
<td>indem</td>
<td>KOUS</td>
<td>indem</td>
</tr>
<tr>
<td>4</td>
<td>bis</td>
<td>KOUS</td>
<td>bis</td>
</tr>
<tr>
<td>4</td>
<td>Daß</td>
<td>KOUS</td>
<td>daß</td>
</tr>
<tr>
<td>3</td>
<td>Ob</td>
<td>KOUS</td>
<td>ob</td>
</tr>
<tr>
<td>2</td>
<td>da</td>
<td>KOUS</td>
<td>da</td>
</tr>
<tr>
<td>2</td>
<td>Solange</td>
<td>KOUS</td>
<td>solange</td>
</tr>
<tr>
<td>2</td>
<td>Sobald</td>
<td>KOUS</td>
<td>sobald</td>
</tr>
<tr>
<td>2</td>
<td>Obwohl</td>
<td>KOUS</td>
<td>obwohl</td>
</tr>
<tr>
<td>2</td>
<td>Damit</td>
<td>KOUS</td>
<td>damit</td>
</tr>
<tr>
<td>1</td>
<td>sooft</td>
<td>KOUS</td>
<td>sooft</td>
</tr>
<tr>
<td>1</td>
<td>sobald</td>
<td>KOUS</td>
<td>sobald</td>
</tr>
<tr>
<td>1</td>
<td>Trotzdem</td>
<td>KOUS</td>
<td>trotzdem</td>
</tr>
<tr>
<td>1</td>
<td>Ehe</td>
<td>KOUS</td>
<td>ehe</td>
</tr>
</tbody>
</table>
“Dass” “that” is also the most frequent subordinating conjunction in the Hfa text, although not to the same extent that it is in Luther, with only 605 occurrences, followed by 252 occurrences of *Wenn* “if/when”.

### 4.3.2 Comparative Frequency of Coordinating Conjunctions

As in the previous chapter, further data analysis takes place at multiple levels, including the same chapters, books, and the entire corpus. The analysis began by looking at the frequency of coordinating conjunctions at the chapter level, followed by the book level, and concluding with the level of the entire corpus. In each case with the coordinating conjunctions, frequency has been calculated both with the conjunction *und* “and” and without.

At the chapter level with *und* “and” included, we can see that Luther consistently has more coordinating conjunctions than Hfa:

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Luther</th>
<th>Hfa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matthew 4</td>
<td>45</td>
<td>32</td>
</tr>
<tr>
<td>Mark 6</td>
<td>155</td>
<td>51</td>
</tr>
<tr>
<td>Luke 15</td>
<td>73</td>
<td>34</td>
</tr>
<tr>
<td>John 1</td>
<td>71</td>
<td>40</td>
</tr>
<tr>
<td>Acts 11</td>
<td>57</td>
<td>22</td>
</tr>
</tbody>
</table>

1 Verses were not included in the analysis for this chapter due to the fact that they often lack any conjunctions.
In each chapter, with *und* “and” included, Luther has considerably more coordinating conjunctions than Hfa, as can be clearly seen in the following figure:

**Figure 4.1**

![Bar Chart](image)

**Figure 4.1**: Coordinating Conjunctions per Chapter

In the example from Mark 6 alone, Luther contains over 100 more examples of coordinating conjunctions than Hfa, indicating the possibility of more clauses in the text. With the number of conjunctions taken as a percentage of all words in the texts, differences between the texts here are found to be significant (*t*(4)=4.62, *p*<.05), using a two-tailed t-test.

Without *und* “and” however, the picture changes considerably in some cases:

**Table 4.6**
As the chart above and the figure below show, without *und* “and”, the frequency of coordinating conjunctions is drastically reduced.

**Figure 4.2**

<table>
<thead>
<tr>
<th>Coordinating Frequency</th>
<th>Luther</th>
<th>Hfa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matthew 4</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Mark 6</td>
<td>20</td>
<td>12</td>
</tr>
<tr>
<td>Luke 15</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>John 1</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>Acts 11</td>
<td>12</td>
<td>5</td>
</tr>
</tbody>
</table>

As the chart above and the figure below show, without *und* “and”, the frequency of coordinating conjunctions is drastically reduced.

**Figure 4.2: Coordinating Conjunctions per Chapter without *und***

In fact, at this level, Hfa actually contains more coordinating conjunctions in two instances: Matthew 4 and John 1. As a result, differences between the two Bible versions are non-significant according to this measure (*t*(4)=1.27 n. s., *p*<.05), once again looking
at the percentage of words that are coordinating conjunctions in relation to the text as a whole.

At the book level, Luther maintains a higher level of coordinating conjunctions in all cases:

Table 4.7

<table>
<thead>
<tr>
<th>Table 4.7: Coordinating Conjunctions per Book</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coordinating Frequency</td>
</tr>
<tr>
<td>-------------------------</td>
</tr>
<tr>
<td>Matthew</td>
</tr>
<tr>
<td>Mark</td>
</tr>
<tr>
<td>Luke</td>
</tr>
<tr>
<td>John</td>
</tr>
<tr>
<td>Acts</td>
</tr>
</tbody>
</table>

In each book, Luther has the greater number of conjunctions by several hundred, if not over a thousand occurrences. Differences between percentages at this level are highly significant ($t(4)=6.31, p<.05$), as depicted graphically in the chart below:

Figure 4.3
With *and* removed from the calculation, the differences between the two Bible versions become less extreme. They are still significant ($t(4)=3.89$, $p<.05$). Unlike the chapter level, however, even without *and*, Luther contains more conjunctions in all books as seen in table 4.8 and figure 4.4:

Table 4.8

<table>
<thead>
<tr>
<th></th>
<th>Coordinating Frequency</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Luther</td>
<td>Hfa</td>
</tr>
<tr>
<td>Matthew</td>
<td>370</td>
<td>291</td>
</tr>
<tr>
<td>Mark</td>
<td>202</td>
<td>188</td>
</tr>
<tr>
<td>Luke</td>
<td>342</td>
<td>330</td>
</tr>
<tr>
<td>John</td>
<td>320</td>
<td>296</td>
</tr>
<tr>
<td>Acts</td>
<td>307</td>
<td>263</td>
</tr>
</tbody>
</table>

Figure 4.4
4.3.3 Comparative Frequencies of Subordinating Conjunctions

The analysis of subordinating conjunctions is similar to that of the coordinating conjunctions, albeit with only one set of calculations. In looking at subordinating conjunctions, the two texts show more similar frequencies, unlike with the overwhelming presence of *und* “and” in Luther. At the chapter level, Luther once again uses more conjunctions than Hfa:

Table 4.9
Table 4.9: Subordinating Conjunctions per Chapter

<table>
<thead>
<tr>
<th></th>
<th>Subordinating Frequency</th>
<th>Hfa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matthew 4</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>Mark 6</td>
<td>20</td>
<td>19</td>
</tr>
<tr>
<td>Luke 15</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td>John 1</td>
<td>15</td>
<td>13</td>
</tr>
<tr>
<td>Acts 11</td>
<td>7</td>
<td>5</td>
</tr>
</tbody>
</table>

In general, the frequencies of subordinating conjunctions are quite low in comparison to coordinating conjunctions. Differences between Luther and Hfa are, however, significant, when looking at subordinating conjunctions as a percentage of all words ($t(4)=2.81, p<.05$). However, in each case, as can be seen in figure 4.5, Luther maintains a higher frequency, sometimes by only one or two occurrences, sometimes by double the number of occurrences:

Figure 4.5
This trend continues at the book level:

Table 4.10

Table 4.10: Subordinating Conjunctions per Book

<table>
<thead>
<tr>
<th></th>
<th>Subordinating Frequency</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Luther</td>
<td>Hfa</td>
</tr>
<tr>
<td>Matthew</td>
<td>340</td>
<td>320</td>
</tr>
<tr>
<td>Mark</td>
<td>277</td>
<td>202</td>
</tr>
<tr>
<td>Luke</td>
<td>413</td>
<td>362</td>
</tr>
<tr>
<td>John</td>
<td>411</td>
<td>351</td>
</tr>
<tr>
<td>Acts</td>
<td>441</td>
<td>344</td>
</tr>
</tbody>
</table>

At the book level, differences range from a low of 20 more occurrences for Luther in the book of Matthew to 97 more occurrences in the book of Acts, compared to the same books in Hfa. Differences in percentages of conjunctions to total words remain significant ($t(4)=3.89$, $p<.05$).

Figure 4.6
In the overall corpus, with regard to coordinating conjunctions, both with *und* “and” and without, and with subordinating conjunctions overall, Luther has the greater number of conjunctions:

Table 4.11

<table>
<thead>
<tr>
<th>Coordinating Frequency</th>
<th>Corpus with <em>und</em></th>
<th>8641</th>
<th>4701</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corpus without <em>und</em></td>
<td>1541</td>
<td></td>
<td>1368</td>
</tr>
<tr>
<td>Subordinating Frequency</td>
<td>1882</td>
<td>1579</td>
<td></td>
</tr>
</tbody>
</table>

When percentages from both corpora at each level of analysis are placed together, the differences between the two texts are significant (*t*(29)=4.11, *p*<.05).

4.3.4 Correlation Analysis of Conjunction Frequency and Lexical Density

In relating these numbers to readability, it is important to know whether there is a correlation between the frequencies of conjunctions and lexical density. As lexical density in this project is calculated as a percentage of lexical words to total words, the comparison here is between lexical density and the percentage of total words that are conjunctions at each level. This is done with lexical density including modals and auxiliaries (high) and lexical density calculated without modal and auxiliaries (low).

Correlation statistics were computed for several categories. In each category, statistics from both texts were combined into one list. This was done in order to see if the entire data set showed a correlation between lexical density and conjunction frequency. The categories are as follow: coordinating conjunctions with *und* “and” and high lexical
density; coordinating conjunctions with \textit{und} “and” and low lexical density; coordinating conjunctions without \textit{und} “and” and high lexical density; coordinating conjunctions without \textit{und} “and” and low lexical density; subordinating conjunctions with high lexical density; subordinating conjunctions with low lexical density. For each case, the values for the two-tailed correlation are as follows: $df = 20; \alpha = 0.05; \text{critical value} = \pm0.423$.

The table below shows the correlation coefficients for each category:

Table 4.12

<table>
<thead>
<tr>
<th>Coordinating</th>
<th>with und</th>
<th>high lexical density</th>
<th>low lexical density</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>without und</td>
<td>high lexical density</td>
<td>low lexical density</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subordinating</td>
<td></td>
<td>high lexical density</td>
<td>low lexical density</td>
</tr>
</tbody>
</table>

In each case, there is a significant negative correlation, indicating that a greater percentage of conjunctions correlates with lower lexical density. These results will be discussed further in the next section.

4.4 Discussion

Looking first at the frequencies of different types of conjunctions used in the two texts, some patterns emerge. The most obvious is the extremely high use of \textit{und} “and” in Luther, compared both with Hfa and with the Luther text’s use of other conjunctions. \textit{Und} “and” outstrips other conjunctions by nearly 7,000 occurrences in Luther alone. In
Hfa, _und_ “and” is once again the most frequent conjunction, although not to the same extent as in Luther. With regard to the use of conjunctions overall, it is clear that in the majority of cases, both with and without _und_ “and”, Luther uses more conjunctions. This is clear in the example below from Matthew 4:6:

**Matthew 4:6 (Luther):** und sprach zu ihm: Bist du Gottes Sohn, _so_ laß dich hinab; _denn_ es steht geschrieben: Er wird seinen Engeln über dir Befehl tun, _und_ sie werden dich auf Händen tragen, auf _daß_ du deinen Fuß nicht an einen Stein stoßest.

**Matthew 4:6 (Hfa):** Spring hinunter! forderte er Jesus auf. Du bist doch Gottes Sohn! _Und_ es steht geschrieben: 'Gott wird seine Engel schicken. Sie werden dich auf Händen tragen, _und_ du wirst dich nicht einmal an einem Stein verletzen!'

The Luther text presents the verse as one long, complex sentence utilizing five conjunctions, as well as punctuation. Hfa, in contrast, uses only two conjunctions and has five sentences. A comparison of the lexical densities of these examples shows that the Luther passage does indeed have a lower lexical density at 45.0, compared to 58.3 for the Hfa passage. Removal of modals and auxiliaries from the lexical density calculation still shows a similar relationship, with Luther at 40.0 and Hfa at 47.2. This is due, in part, to the clausal complexity of the Luther passage.

Halliday and Matthiessen (2004) relate lexical density and clausal complexity in the following way:

> Typically, written language becomes complex by being lexically dense: it packs a large number of lexical items into each clause; whereas spoken language becomes complex by being grammatically intricate: it builds up elaborate clause complexes out of parataxis and hypotaxis. (p. 654)

As the sections compared in Luther and Hfa are from identical passages, they should carry the same message. As previously stated, Luther has consistently lower lexical
density. Furthermore, Luther seems to be more grammatically intricate, due to the higher number of both coordinating and subordinating conjunctions. This, as Halliday and Matthiessen stated above, is a hallmark of *spoken* intricacy. It is also clear from the correlation analyses that the percentage of conjunctions used in a text has a bearing on lexical density.

Lower use of conjunctions in Hfa as compared to Luther can also be attested to as a historical development of sentence structure in German. According to König (1998):


Statistical examinations have shown, that sentence length has decreased in comparison to the 18th and 19th centuries. In aesthetic literature, sentences were, at that time, most often 25-35 words in length (Lessing 24, Goethe 30), today in this area, there are sentences with 10-12 words. (translation mine).

This decreased sentence length seems to have the effect of packing the information of the passage more densely. Use of more clauses spreads the information, adding more grammatical words and lowering the overall concentration of lexical words.

### 4.5 Conclusion

It is clear from the frequency comparisons and the correlations that conjunctions have an important relationship to lexical density, and therefore to readability. The greater use by Luther of both coordinating and subordinating conjunctions lends further evidence to the assertion that Luther was more readable for its contemporaries than Hfa is for its
respective contemporaries. The greater use of conjunctions denotes greater clausal complexity and lower lexical density. Further study in this area would be greatly aided by a disambiguation of such conjunctions as *und* “and” and *oder* “or”, which can have roles other than that of clausal conjunctions. However, it is clear from the statistical analysis done in this chapter that greater use of conjunctions does promote lower lexical density.
Chapter 5

Nominalization

5.1 Introduction

This chapter examines the presence of nominalization in the two Bible texts and explores their effects on readability. To reiterate from section 2.6, Halliday and Matthiessen (2004, p. 636) describe nominalization as a form of grammatical metaphor, the presence of which can make a text more challenging. In such nominalizations, “processes and qualities are construed as if they were entities” (p. 637). According to Taverniers (2003), written language has more such metaphors than spoken language, which is “attributed to a more general difference in types of complexity: written language is said to be ‘lexically dense’, whereas spoken language is ‘grammatically intricate’” (p. 9).

According to Unsworth (1999), the use of nominalizations has led to difficulties in the processing of scientific and even historical texts by students. Specifically, “for students who are confident users of spoken English but lack familiarity with the grammar of the written form, the greater lexical density of school texts can contribute to comprehension difficulties” (p. 510). The increase in the use of nominalizations in such texts is undoubtedly a contributing factor to their difficulty. The ability to construct and process the specialized knowledge of scientific texts is dependent on a reader’s ability to process such forms (p. 511).
It is the primary argument of this chapter that the same holds true for the processing of nominalizations in the Bible texts. The issue of nominalization will be examined in two ways in this chapter: 1) via statistical means through corpus analysis, and 2) in context, looking at highly frequent nominalizations and how they are unpacked in the text, if at all. The frequency analysis will be based on corpus searches focusing on specific suffixes attributed to nominalizations. The contextual analysis that follows will look at highly frequent nominalizations containing the suffix –ung.

5.2 Methodology

As stated above, two means of analysis are employed in the exploration of nominalization in this chapter. The first involves corpus-based statistical analyses for the numbers of nominalizations found in the text. The second analysis takes types of nominalizations found in the statistical analyses and explores them in context. The next two sections will describe specifically how each analysis was conducted.

5.2.1 Statistical Analysis

As with previous chapters, the corpora for this analysis are the Gospel books from Luther and Hfa. This includes the books Matthew, Mark, Luke, John, and the Acts of the Apostles. This grouping of books was chosen for their similar nature in terms of genre. These sections are narrative in nature, where other sections of the New Testament show a more epistolary nature. Both corpora are of very similar size (Luther: 101,629 tokens;
Hfa: 102,610 tokens). The analysis is based primarily on a frequency list made from both texts.

Statistical analyses of the frequency results of this analysis are done using the log-likelihood ratio test. Because of the relatively small size of the corpora, which could often result in low frequency counts, this method is preferable to the chi-squared test (Rayson, et al., 2004, p.4). This test takes as its null hypothesis the assumption that there will be no significant difference between the two corpora based on frequency and corpus size. The null hypothesis is rejected if significant results are returned, much like a chi-squared test (p. 3).

While null suffix nominals are often difficult to search for, simple UNIX tools can easily find others within a corpus, such as those ending in \(-ung, \-heit, \-keit, \ or \-nis\). Both \-ung and \-heit fall into the top seven cited by Russ, as can be seen above. Using a frequency list compiled from the two corpora, the frequency of nominalizations formed with the above suffixes was found using the \texttt{grep} function in UNIX. \texttt{Grep} is a tool that searches for patterns within lines of text (Brew and Moens, 2002, p. 36). The details of this process can be seen in Appendix C. The results are the basis for the statistical analysis in this chapter.
5.2.2 Analysis of Nominalizations in Context

The analysis of nominalizations in context follows the work of Unsworth (1999). The following table from that work shows a brief analysis of how nominalizations are unpacked in a secondary school history text:

Figure 5.1

<table>
<thead>
<tr>
<th>Nominalization</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>help pay</td>
<td>Process (verb)</td>
</tr>
<tr>
<td>migrate</td>
<td>Process (verb)</td>
</tr>
<tr>
<td>assisted migration</td>
<td>Nominal group Classifier + Thing</td>
</tr>
</tbody>
</table>

more people to come and live in the colonies. (Anderson & Ashton, 1993, p. 135)

When the war ended about 595,000 Australians were serving in the forces. As a general rule they were given their discharges under a points system, whereby age, length of service and family were taken into account. The demobilization was effected quickly and smoothly as dispersal centres were established in each State. (Simmelhaig & Spenceley, 1984, p. 126)

Figure 5.1: Use of Nominalization in Secondary School History Texts (Unsworth 1999)

Nominalizations within the Bible texts were approached in a similar way. A sample of examples from the most frequent –ung nominalizations were examined in the context of their chapters to see if or how these nominalizations are unpacked within the text. These nominalizations were chosen because of the types examined in this chapter, they are the most frequent.
Essentially, this section involves a search for the congruent forms of the most frequent types of -ung nominalizations. According to Demske (2000), “Die Geschichte der ung-Nominalisierung seit dem 17. Jh. ist eine Geschichte abnehmender Produktivität” (The history of the ung-nominalization since the 17th century is a history of decreasing productivity) (p.394, translation mine). It seems then, that these nominalizations have, over time come to fulfill specific purposes. Demske goes on to say that, in their unmarked form, -ung nominalizations are based on telic verbs in present-day German that deal with experiences and resulting conditions (p. 398). Telicity indicates that a process is complete, and there are examples in this analysis of nominalizations of this type, such as Versuchung, which means the experience of temptation, and Auferstehung, the condition found after rising from death.

It is important to understand, especially in the context of readability, what the alternatives to nominalization are. Heyvaert (2003) states that “the systemic-functional perspective on nominalizations is closely tied up with the concept of grammatical metaphor: nominalization is presented as a major resource for the creation of ‘metaphorical’ rather than ‘typical’ or ‘congruent’ lexicogrammatical realizations of semantic categories” (p. 65). The author gives the following example of a grammatically metaphoric structure and its congruent counterpart:

The cast’s brilliant acting – grammatical metaphor

The cast acted brilliantly – congruent form (ibid.)

The congruent form is known as the metaphor’s “agnate” (ibid.).

Analyses of such agnate forms, according to Heyvaert (2003), have the following tendencies:
- Emphasis is put on finding one congruent structure as the agnate for a nominalization, and that must be grammatically well formed itself. Example: her sailing out of the room/she sailed out of the room.

- The congruent agnate of a deverbal nominalization (as opposed to a deadjectival nominalization) is typically clausal. Example: the allocation of an extra packer/they allocate an extra packer (p. 66)

Heyvaert argues that “each nominalization should be viewed as the metaphorical counterpart of not one congruent agnate, but of a network of agnate structures, clausal and other” (p. 67). Such a view of nominalization, Heyvaert argues, opens up the possibility for studying the relationship between nominalizations and clausal forms, as well as looking at nominalization and metaphor as forming “a unique combination of linguistic categories” (p. 95).

### 5.3 Data Analysis

#### 5.3.1 Statistical Analysis of Nominalization Types

The tables below look at different aspects of the four types of nominalization mentioned above. The first four tables present the ten most frequent types from each set. The last table looks at an overall comparison of occurrences of these types. Of the types, -ung nominalizations are of the deverbal variety, and the others are all deadjectival. In each table, types that appear in both lists of the ten most frequent nominalizations are highlighted in yellow.

Table 5.1
Of the types of –ung nominalizations that overlap, none show significant differences when examined using the log-likelihood test (Auferstehung: \( LL=0.25, p<.05, \text{NS} \); Vergebung: \( LL=0.00, p<.05, \text{NS} \); Gotteslästerung: \( LL=1.11, p<.05, \text{NS} \); Versuchung: \( LL=0.10, p<.05, \text{NS} \)) It is interesting to note, however, that only four of the top ten overlap.

Table 5.2

Table 5.1: Ten most frequent –ung nominalizations

<table>
<thead>
<tr>
<th>rank</th>
<th>tokens</th>
<th>type</th>
<th>rank</th>
<th>tokens</th>
<th>type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25</td>
<td>Auferstehung</td>
<td>1</td>
<td>29</td>
<td>Auferstehung</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>Vergebung</td>
<td>2</td>
<td>11</td>
<td>Vergebung</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>Verheissung</td>
<td>3</td>
<td>10</td>
<td>Meinung</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>Hoffnung</td>
<td>4</td>
<td>9</td>
<td>Kleidung</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>Verwesung</td>
<td>5</td>
<td>9</td>
<td>Heilung</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>Reinigung</td>
<td>6</td>
<td>9</td>
<td>Gotteslästerung</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>Nahrung</td>
<td>7</td>
<td>8</td>
<td>Wohnung</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>Gotteslästerung</td>
<td>8</td>
<td>6</td>
<td>Festung</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>Anfechtung</td>
<td>9</td>
<td>6</td>
<td>Erfuellung</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>Versuchung</td>
<td>10</td>
<td>5</td>
<td>Versuchungen</td>
</tr>
</tbody>
</table>

Table 5.2: Ten most frequent –heit nominalizations

<table>
<thead>
<tr>
<th>rank</th>
<th>tokens</th>
<th>type</th>
<th>rank</th>
<th>tokens</th>
<th>type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>29</td>
<td>Wahrheit</td>
<td>1</td>
<td>40</td>
<td>Wahrheit</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>Weisheit</td>
<td>2</td>
<td>15</td>
<td>Gelegenheit</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>Gewohnheit</td>
<td>3</td>
<td>9</td>
<td>Weisheit</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>Krankheit</td>
<td>4</td>
<td>8</td>
<td>Krankheiten</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>Klarheit</td>
<td>5</td>
<td>6</td>
<td>Krankheit</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>Schalkheit</td>
<td>6</td>
<td>5</td>
<td>Bosheit</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>Bosheit</td>
<td>7</td>
<td>4</td>
<td>Freiheitskämpfer</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>Krankheiten</td>
<td>8</td>
<td>4</td>
<td>Dunkelheit</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>Gesundheit</td>
<td>9</td>
<td>3</td>
<td>Sicherheit</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>Unwissenheit</td>
<td>10</td>
<td>3</td>
<td>Angelegenheit</td>
</tr>
</tbody>
</table>
Overlaps in the –heit category show the same results as the –ung nominalizations, with no significant differences (Wahrheit: $LL=1.62, p<.05$, NS; Weisheit: $LL=1.16, p<.05$, NS; Krankheit: $LL=1.03, p<.05$, NS; Bosheit: $LL=0.10, p<.05$, NS), and the same can be seen for the similar types of –keit nominalizations below (Herrlichkeit: $LL=0.00, p<.05$, NS; Gerechtigkeit: $LL=1.56, p<.05$, NS; Barmherzigkeit: $LL=1.06, p<.05$, NS).

Table 5.3

<table>
<thead>
<tr>
<th>keit from Luther</th>
<th>rank</th>
<th>tokens</th>
<th>type</th>
<th></th>
<th>rank</th>
<th>tokens</th>
<th>type</th>
</tr>
</thead>
<tbody>
<tr>
<td>&amp;1</td>
<td>26</td>
<td>Herrlichkeit</td>
<td>1</td>
<td>26</td>
<td>Herrlichkeit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>Gerechtigkeit</td>
<td>2</td>
<td>14</td>
<td>Öffentlichkeit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>Barmherzigkeit</td>
<td>3</td>
<td>8</td>
<td>Schwierigkeiten</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>Obrigkeit</td>
<td>4</td>
<td>7</td>
<td>Wirklichkeit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>Traurigkeit</td>
<td>5</td>
<td>6</td>
<td>Gerechtigkeit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>Freudigkeit</td>
<td>6</td>
<td>6</td>
<td>Barmherzigkeit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>Ungerechtigkeit</td>
<td>7</td>
<td>3</td>
<td>Deutlichkeit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>Härteigke</td>
<td>8</td>
<td>2</td>
<td>Überheblichkeit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>Ewigkeit</td>
<td>9</td>
<td>2</td>
<td>Möglichkeit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>Niedrigkeit</td>
<td>10</td>
<td>2</td>
<td>Einsamkeit</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The –nis category, however, does show some significant differences:

Table 5.4
The term *Gleichnis*, which is often seen in English as “parable,” appears significantly more often in Luther than in Hfa \((LL = 8.88; p < 0.01, \text{ critical value } = 6.63)\). The plural form *Gleichnisse* also showed significant differences, once again with the higher frequency in Luther \((LL = 4.94; p < 0.05, \text{ critical value } = 3.84)\).

While the ten most frequent members of each group may not show conclusive differences between the texts, a look at all types of each form shows much different results. When all forms are taken into account, Hfa tends to show a significantly greater use of these nominalizations. This is clearly shown in the table below:

**Table 5.5**

<table>
<thead>
<tr>
<th>Suffix</th>
<th>Luther</th>
<th>Hfa</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>total types</td>
<td>total tokens</td>
</tr>
<tr>
<td>-ung</td>
<td>47</td>
<td>144</td>
</tr>
<tr>
<td>-heit</td>
<td>18</td>
<td>89</td>
</tr>
<tr>
<td>-keit</td>
<td>17</td>
<td>78</td>
</tr>
<tr>
<td>-nis</td>
<td>20</td>
<td>172</td>
</tr>
</tbody>
</table>
Using the log-likelihood calculation, significant differences were found for each of the above examples, except for –keit (for –ung $LL=101.41; p<0.0001$, critical value=15.13; for –heit $LL=4.06; p<0.05$, critical value=3.84; for -nis $LL=14.86; p<0.001$, critical value=10.83).

5.3.2 Analysis of Nominalizations in Context

In exploring the agnate patterns in both Luther and Hfa, interesting and unexpected patterns were found. The following sections examine the most frequent nominalizations ending in –ung.

5.3.2.1 Auferstehung

Focusing on the most frequent –ung nominalization Auferstehung “resurrection”, it appears that both versions used agnate forms inconsistently in close context with the nominalization. This can be most clearly seen in a passage that appears in the books of Matthew, Mark and Luke, in which Auferstehung is used multiple times. Below are examples from Luther and Hfa of the passage in Matthew:

Matthew 22: 23-32 (Luther): 23 An dem Tage traten zu ihm die Sadduzäer, die da halten, es sei kein Auferstehen, und fragten ihn 24 und sprachen: Meister, Mose hat gesagt: So einer stirbt und hat nicht Kinder, so soll sein Bruder sein Weib freien und seinem Bruder Samen erwecken. 25 Nun sind bei uns gewesen sieben Brüder. Der erste freite und starb; und dieweil er nicht Samen hatte, ließ er sein Weib seinem Bruder; 26 desgleichen der andere und der dritte bis an den siebenten. 27 Zuletzt nach allen starb auch das Weib. 28 Nun in der Auferstehung, wes Weib wird sie sein unter den sieben? Sie haben sie ja alle gehabt. 29 Jesus aber
antwortete und sprach zu ihnen: Ihr irrt und wisset die Schrift nicht, noch die Kraft Gottes. 30 In der Auferstehung werden sie weder freien noch sich freien lassen, sondern sie sind gleichwie die Engel Gottes im Himmel. 31 Habt ihr nicht gelesen von der Toten Auferstehung, was euch gesagt ist von Gott, der da spricht: 32 "Ich bin der Gott Abrahams und der Gott Isaaks und der Gott Jakobs"? Gott aber ist nicht ein Gott der Toten, sondern der Lebendigen.

In this first example from Luther, the nominalization appears three times, as well as a slightly different form of the nominalization in verse 23, which uses an –en ending as opposed to the –ung ending. At no point in the passage, however, do we see a clausal agnate of Auferstehung “resurrection”, namely auferstehen “to resurrect”. The situation is similar in the passage from Hfa:


This time there are only three appearances of the nominalization. There is, however, one appearance of the agnate verb form auferstehen “to resurrect”.

Other passages in Luther, however, do contain agnate forms of the nominalization. The first examples also come from the book of Matthew:
Matthew 27: 52-53 (Luther): 52 Und die Erde erbebte, und die Felsen zerrissen, die Gräber taten sich auf, und **standen auf viele Leiber** der Heiligen, die da schliefen, 53 und gingen aus den Gräbern nach seiner **Auferstehung** und kamen in die heilige Stadt und erschienen vielen.

In this example from Luther, the agnate can be seen preceding the nominalization. In this case, **Auferstehung** “resurrection” finds its congruent form in **standen auf viele Leiber** “many bodies awoke”. This agnate contains the verbal form of the nominalization, simply with the prefix **auf** unattached. The agnate from Hfa is similar, although it uses a different verb:

Matthew 27: 52-53 (Hfa): 52 Gräber öffneten sich, und viele Verstorbene, die nach Gottes Willen gelebt hatten, **erwachten vom Tod**.
53 Nach der **Auferstehung** Jesu verließen sie ihre Gräber, gingen in die Stadt und erschienen dort vielen Leuten.

In this case from Hfa we have a clausal agnate form – **erwachten vom Tod** “awoke from death”. This gives the reader a clear idea of what the nominalization **Auferstehung** “resurrection” means. Further clarification of the term is given in both of the following examples:

Acts 2: 30-31 (Luther): 30 Da er nun ein Prophet war und wußte, daß ihm Gott verheißen hatte mit einem Eide, daß die Frucht seiner Lenden sollte auf seinem Stuhl sitzen, 31 hat er's zuvor gesehen und geredet von der **Auferstehung** Christi, daß **seine Seele nicht dem Tode gelassen ist** und sein Fleisch die Verwesung nicht gesehen hat.


In the example from Luther, **Auferstehung** “resurrection” has the agnate for **seine Seele nicht dem Tode gelassen ist** “his soul is not left to the dead”. Similarly, the agnate form in Hfa is **Er wird nicht bei den Toten bleiben** “he will not remain with the dead”.
5.3.2.2 *Gotteslästerung*

Another frequent nominalization that appears in both lists is *Gotteslästerung* “blasphemy”. Much like *Auferstehung* “resurrection”, there are few examples where the text employs both the nominalization and its agnate form. Below we see a situation where this does take place to some extent in Luther, but not in Hfa:

**Matthew 26:65 (Luther):** 65 Da zerriß der Hohepriester seine Kleider und sprach: *Er hat Gott gelästert!* Was bedürfen wir weiteres Zeugnis? Siehe, jetzt habt ihr seine *Gotteslästerung* gehört.

**Matthew 26:65 (Hfa):** 65 Empört zerriss der Hohepriester sein Gewand und rief: »Das ist *Gotteslästerung*! Wozu brauchen wir noch weitere Zeugen? Ihr habt seine *Gotteslästerung* ja selbst gehört!

In the passage from Luther, we first see the agnate form using the verbal form of the nominalization, *Er hat Gott gelästert* “He has blasphemed God”. This is then followed by the nominalization itself. In the parallel verse from Hfa, the nominalization is simply used twice, with no agnate.

5.3.2.3 *Versuchung*

The final nominalization that appears in both texts, and that will be discussed in context, is *Versuchung* “temptation”. There are only four occurrences of this nominalization in Luther, all in the singular form, and seven in Hfa, including both singular and plural. Hfa uses the nominalization in three of the four verses that Luther does, an example of which can be seen below:

**Matthew 6:13 (Luther):** 13 Und führe uns nicht in *Versuchung*, sondern erlöse uns von dem Übel. Denn dein ist das Reich und die Kraft und die Herrlichkeit in Ewigkeit. Amen.

The passage is one that would be familiar to all Christians as being part of the Lord’s Prayer. Both texts use the same nominalization with no agnate.

In one case where Luther uses Versuchung, Hfa does not, and this can be seen below:


Where Luther simply uses Versuchung “temptation”, Hfa employs a more congruent form of the process, Jesus weiter auf die Probe zu stellen “to test Jesus further”, expanding what is one word in Luther to an entire clause.

In one case, Hfa utilizes an agnate form in the verse along with the nominalization:

Matthew 18:7 (Hfa): 7 Wehe der Welt, denn sie verführt zum Unglauben! Solche Versuchungen können ja nicht ausbleiben. Aber wehe dem, der daran schuld ist!

The agnate form precedes the nominalization in this example, explaining the type of temptation taking place: verführt zum Unglauben (lead astray to unbelief). The parallel verse from Luther does not use the nominalization or an agnate form. Rather, as can be seen below, the focus is on three occurrences of the term Ärgernis (upset, annoyance), a deadjectival nominalization with the suffix -nis:

Matthew 18:7 (Luther): 7 Weh der Welt der Ärgernisse halben! Es muß ja Ärgernis kommen; doch weh dem Menschen, durch welchen Ärgernis kommt!
5.3.2.4 Verheißung

The nominalization *Verheißung* “promise” is the first nominalization to be examined that does not appear in both texts, occurring only in the Luther text. The term itself means promise, and its use in Luther can be seen in the example below:

**Luke 24:49 (Luther):** 49 Und siehe, ich will auf euch senden die **Verheißung** meines Vaters. Ihr aber sollt in der Stadt Jerusalem bleiben, bis ihr angetan werdet mit der Kraft aus der Höhe.

In the above verse it is not made clear exactly what the promise to be sent is. The same verse in Hfa, however, uses an agnate form of **Verheißung**, and is explicit about what is being sent:

**Luke 24:49 (Hfa):** 49 Ich werde euch den **Heiligen Geist** geben, **den mein Vater euch versprochen hat**. Bleibt hier in Jerusalem, bis ihr diese Kraft von oben empfangen habt!«

The text in Hfa states clearly that it is the *Heiligen Geist* “Holy Ghost” that was promised, and this is expressed in the clausal form *den mein Vater euch versprochen hat* “that my Father promised to you”.

A similar comparison can be made with the following parallel verses from the book of Acts:

**Acts 1:4 (Luther):** 4 Und als er sie versammelt hatte, befahl er ihnen, daß sie nicht von Jerusalem wichen, sondern warteten auf die **Verheißung** des Vaters, welche ihr habt gehört (sprach er) von mir;

**Acts 1:4 (Hfa):** 4 Als sie an einem dieser Tage miteinander aßen, sagte Jesus zu seinen Jüngern: »Verlasst Jerusalem nicht! Bleibt so lange hier, bis in Erfüllung gegangen ist, **was euch der Vater durch mich versprochen hat**.

Once again, the nominalization in Luther is replaced by an entire clause in Hfa. It is not always the cause that *Verheißung* is replaced by a complete clause, however. There are
occurrences where the nominalization found in Luther is simply replaced with another – *en* nominalization, as seen in these examples:

**Acts 7:17 (Luther):** 17 Da nun die Zeit der Verheißung nahte, die Gott Abraham geschworen hatte, wuchs das Volk und mehrte sich in Ägypten,

**Acts 7:17 (Luther):** 17 Dann kam die Zeit, in der Gott das Versprechen erfüllen wollte, das er Abraham gegeben hatte. Die Nachkommen Josefs und seiner Brüder waren in Ägypten zu einem großen Volk geworden.

In this case, the nominalization *Verheißung* “promise” is replaced by another nominalization that translates as promise, in this case *Versprechen*.

### 5.3.2.5 Anfechtung

*Anfechtung* “attack” presents another example of a nominalization that appears only in Luther. In this case however, its equivalent in parallel verses in Hfa is not an agnate form. Rather it is often another nominalization:

**Matthew 26:41 (Luther):** 41 Wachet und betet, daß ihr nicht in Anfechtung fallet! Der Geist ist willig; aber das Fleisch ist schwach.

**Matthew 26:41 (Hfa):** 41 Bleibt wach und betet, damit ihr der Versuchung widerstehen könnt. Ich weiß, ihr wollt das Beste, aber aus eigener Kraft könnt ihr es nicht erreichen.«

This first set of parallel verses establishes the pattern – where Luther uses *Anfechtung*, implying some sort of attack or impugning, Hfa uses the nominalization *Versuchung* “temptation”, which we have already discussed above. A very similar pairing occurs again in the book of Luke:

**Luke 22:40 (Luther):** 40 Und als er dahin kam, sprach er zu ihnen: Betet, auf daß ihr nicht in Anfechtung fallet!

In yet another example, where Luther uses Anfechtung, Hfa uses not only Versuchung “temptation”, but also Gefahr “danger”:

Luke 22:28 (Luther): 28 Ihr aber seid's, die ihr beharrt habt bei mir in meinen Anfechtungen.


Finally, in this example we see the Hfa equivalent to Anfechtung “attack” presented as Schwierigkeiten “difficulties”:

Luke 8:13 (Luther): 13 Die aber auf dem Fels sind die: wenn sie es hören, nehmen sie das Wort mit Freuden an; und die haben nicht Wurzel; eine Zeitlang glauben sie, und zur Zeit der Anfechtung fallen sie ab.


Clearly, in all the examples above there are patterns of interest to be discussed, both from the statistical data and from the examples of nominalization in context. The next section will discuss both of these areas in relation to readability.

5.4 Discussion

5.4.1 Comparison of the Most Frequent Nominalizations in Luther and Hfa

It is possible that greater use of nominalizations could make Hfa more difficult in some respects than Luther. The greater frequency of nominalizations is consistent with
Hfa’s typically higher concentration of noun usage, as seen in chapter 3. One must also, however, take the reader into account. One means of doing so is to look at the results of the most frequent overlaps in each group of nominalizations. In all cases except for –nis, the most frequent nominalizations in each group match up:

- *ung*: Auferstehung
- *heit*: Wahrheit
- *keit*: Herrlichkeit

Each of these terms has great biblical significance: Auferstehung “resurrection” refers to a key tenet of Christianity, the resurrection of Jesus Christ; Wahrheit, or “truth”, is a difficult term both religiously and philosophically; and Herrlichkeit “glory” often refers to the glory of God. If we look at these words in terms of a possible network of agnates, all of them could have extensive agnate networks. In terms of a specific type of reader, i.e. a reader new to the Bible, one that might typically look first at Hfa, access to the necessary agnate networks could very well be limited. This brings us to the non-significant differences in frequency between the two versions. One would think that a version that seeks to reach those unfamiliar with the Bible and Christianity would seek to minimize such difficult terms. Instead, it seems that little or no effort was made to reduce their appearance in Hfa. Not only are these highly frequent nominalizations not reduced in Hfa, more with the same suffixes are found there than are found in Luther. One could question whether or not it is more helpful for a new reader of the Bible to have such terms present, or to have them broken down into a more congruent agnate. While Auferstehung will be discussed later, it is helpful here to compare occurrences of Wahrheit and Herrlichkeit in context:
John 3:21 (Luther): Wer aber die Wahrheit tut, der kommt an das Licht, daß seine Werke offenbar werden; denn sie sind in Gott getan.

John 3:21 (Hfa): Wer aber die Wahrheit Gottes liebt und das tut, was er will, der tritt ins Licht! An ihm zeigt sich: Gott selber bestimmt sein Handeln.

In neither case above is Wahrheit clarified with an agnate form.

With Herrlichkeit, we see a similar example from Luther, with a slight difference, however, in Hfa:

Matthew 6:29 (Luther): Ich sage euch, daß auch Salomo in aller seiner Herrlichkeit nicht bekleidet gewesen ist wie derselben eins.

Matthew 6:29 (Hfa): Ich sage euch, selbst König Salomo war in seiner ganzen Herrlichkeit nicht so prächtig gekleidet wie eine dieser Blumen.

In the Luther version we see no agnate form of Herrlichkeit. Hfa provides more information to the reader through the adjective phrase prächtig gekleidet “gloriously dressed”. In this way, the reader is made more aware that the glory in question is related to fine clothing.

The question of why some cases of the terms above have agnates in Hfa and some do not could be a question of frequency within the current German language. An examination of a frequency list of German word forms from the Leeds Internet Corpus shows that Wahrheit “truth” has a frequency of 88.21 instances per million words (ipm) within a corpus containing 187,789,449 tokens and 4,394,346 types (Rapp, http://corpus.leeds.ac.uk/frqc/internet-de-forms.num, last accessed on 5-15-2008). Herrlichkeit, in contrast, only has a frequency of 4.04 ipm. For comparison, the most frequent noun in the list, Menschen “people”, has a frequency of 773.21 ipm. Wahrheit “truth”, with its relatively higher frequency is without agnate form in the examples
above. *Herrlichkeit* “glory” has considerably lower frequency, and does also have some clarification in Hfa. Although such a conclusion must remain speculative at this point, it is conceivable that relative frequencies of nominalizations influence whether or not a translator would include an agnate form or other clarification in the text. Higher frequency could lead to the assumption of familiarity, whereas lower frequency could indicate that a term is less well known and more in need of clarification. This will be discussed more in the sections below.

### 5.4.2 Comparison of Nominalizations in Context

#### 5.4.2.1 Overlapping Nominalizations

In the cases of overlapping nominalizations, that is, those that appear on both frequency lists, there does not seem to be any consistent pattern in which one text couples agnate usage with nominalization more than another. From the examples seen above, both texts have examples of nominalization with agnates present nearby, and examples where they are absent, such as the first Luther example of *Auferstehung* in Matthew 22:23-32 and the Hfa example of *Gotteslästerung* in Matthew 26:65. In looking back at the log-likelihood significance tests from section 5.3.1, neither nominalization showed a significantly more likely chance of appearing in one Bible version or the other. In both cases, the reader has to rely upon the linguistic and stylistic choices of the translator as far as whether or not nominalizations are clarified in context.
In this case, it is left once again to the frequency of nominalizations overall to determine which text is more readable from the standpoint of nominalization use. In the case of all nominalizations in the –ung category, Hfa shows more frequent use. Use overall, however, is relatively low, with even the most frequent –ung nominalization in Hfa appearing only 29 times, out of a total of 102,610 tokens in the text. When looking at the frequency of these terms in the Leeds Corpus, Auferstehung “resurrection”, the most frequent nominalization in the two versions, has a frequency of 5.96 ipm, ranging to no occurrences of Gotteslästerung “blasphemy”. Looking even at other nominalizations such as Wahrheit “truth” that appear in both texts, these words are relatively infrequent in modern German. One would expect to see more clarification then within the Hfa text, if it is trying to be more readable than Luther.

The question here becomes, does this relatively low frequency of nominalizations in both texts serve to make such nominalizations more difficult for readers? If that is the case, it can be said that neither text really aids the reader more than the other, at least in the case of overlapping –ung nominalizations.

5.4.2.2 Non-overlapping Nominalizations

Examining non-overlapping nominalizations shows a somewhat different pattern. The examples of Anfechtung “attack” and Verheißung “promise”, while sometimes replaced by agnate forms, are also replaced with other nominalizations. In the example of Anfechtung in Luke 22:40, Hfa replaced the nominalization with another nominalization found in both texts – Versuchung “temptation”. Versuchung “temptation” has a
frequency of 6.50 ipm according to the Leeds corpus (Rapp, http://corpus.leeds.ac.uk/frqc/internet-de-forms.num, last accessed on 5-15-2008). While it may seem that the frequency of Versuchung “temptation” is quite low in this corpus, it must be realized that Anfechtung “attack” does not appear in the corpus at all. It appears then, that while not using an agnate form to replace Anfechtung “attack”, the translators of Hfa opted for a more frequently used nominalization. This is further evidenced from the Leeds corpus when one examines the other example of replacement with an –en nominalization above – Schwierigkeiten “difficulties” in Luke 8:13. Schwierigkeiten “difficulties” is far more frequent in the Leeds corpus, at 57.79 ipm. The same is true for the examples with Verheißung “promise”. While this term does appear in the corpus with a frequency of 2.31 ipm, the nominalization used to replace it in the Hfa version of Acts 7:17 – Versprechen “promise” – has a frequency of 13.73 ipm. In the other cases with Anfechtung “attack”, we do see it replaced with an agnate form, making Hfa easier to read in these examples.

The issue of the replacement of non-overlapping nominalizations could be extended to other content words in general, the question being, are there highly frequent content words in Luther that are, like Anfechtung “attack” and Verheißung “promise”, not used in Hfa? A comparison of the 20 most frequent nouns, verbs and adjectives in the two texts shows that, rather than having very different vocabulary, the two texts use very similar terms. Of the twenty most frequent nouns in Luther, only three are not found in some form in the list from Hfa: Propheten “prophets”, Namen “names”, and Volk “people/nation”. Propheten appears 97 times in both Luther and Hfa. Namen “names” shows greater difference, appearing 124 times in Luther and 41 times in Hfa. Volk
“people/nation” appears 235 times in Luther, compared to 102 times in Hfa. While these terms are less frequent in Hfa, they are not unused, as seen with Anfechtung “attack”. Of the verbs, only two from Luther do not appear among the 20 most frequent in Hfa: ward “became” and stand “stood”, which appear in Luther 186 and 77 times respectively. In comparison, they appear in Hfa 0 times for ward and 53 times for stand. Ward, however, is a form of werden, and has been replaced by the modern form wurde in current usage. Werden means both “to become” and is also the marker for future tense, and thus very frequent in both texts in different forms. Finally in the adjectives, only one from Luther does not appear at all in Hfa, namely selig “blessed”, while it appears 34 times in Luther. Thus, the elimination of terms seems to be limited, in most cases, to fairly infrequent terms, like Anfechtung “attack”, which, even in Luther, only appears five times.

The most telling data come from the frequency comparison of all types of –ung nominalizations and the other types examined as well, as seen at the end of section 5.3.1. Looking at the four derivational suffixes, Hfa has significantly more nominalizations. However, there is some variation depending upon the particular suffix. The frequency of nominalizations using –keit is not significantly different, and Luther contains significantly more nominalizations using the relatively infrequent –nis. In sum, the overall trend toward increased use of nominalization seems clear. Diachronic differences in the productivity of each suffix may account for the anomalous findings for –nis and –keit, but this requires further research. Since neither text seems to make a concentrated effort to clarify nominalizations, the evidence that Luther contains significantly fewer occurrences of these nominalizations lends support to the conclusion that Luther was a
more readable text for its contemporaries in this respect, since it subjected its readers to lower numbers of difficult terms.

5.5 Conclusion

In looking at the examples of overlapping and non-overlapping nominalizations, the question then becomes, why were some nominalizations replaced in Hfa and others not? Frequency seems to be part of the answer. Most of the examples used in Hfa appear on the Leeds frequency list, with the exception of Gotteslästerung. From a frequency standpoint then, it seems that Hfa occasionally eliminates nominalizations (and assumedly other word types) and replace them with more frequently used equivalents, which is consistent with the Dynamic Equivalence method of translation (Nida, 1964). Thus, while not necessarily lessening the difficulty of nominalizations via the use of agnate forms, the reader is aided by the fact that those nominalizations that do appear are more frequent and more familiar.

When looking at readability from the perspective of nominalization in these texts, and with –ung nominalizations in particular, we find that there is a difficult balance here. On the one hand, lesser use of nominalization should give the reader greater ease of reading, since he or she would encounter these supposedly more difficult forms less often. On the other hand, low frequency terms also seem to increase difficulty, especially when not coupled with agnate forms. As stated at the beginning of the discussion section, it is the reader’s knowledge and familiarity with the terms that may further determine the readability of these texts.
Chapter 6
Lexical Bundles

6.1 Introduction

As discussed in section 2.7, lexical bundles are “sequences of three or more words that statistically co-occur in a register” (Cortes, 2004, p. 400). These bundles are important parts of language. Biber and Conrad (1999) state: “Given the high frequency of lexical bundles, it is reasonable to assume that they are basic building blocks for both spoken and written discourse” (p. 188). The goal of this section is two-fold: first, the texts will be examined for statistical differences in the number and types of lexical bundles found; second, the types of lexical bundles will be compared in context, looking at the roles played by the most prevalent bundles found in each text. The discussion in this chapter will focus on how the statistical and contextual differences enhance or detract from the texts’ readability, emphasizing the following question: based on the concentration and use of lexical bundles, which text is a more readable Bible for its contemporary readers?

Such a study of lexical bundles is important for two reasons. Greater use of lexical bundles in a text should make a text more readable. According to Biber and Conrad (1999), lexical bundles encompass often very complex grammatical structures (p. 188), notwithstanding the fact that they are so frequent. In a language learning situation, despite such complexity, frequent bundles are often among the earliest units learned by
students (ibid.). This brings us then to readability in relation to the frequency of these bundles. By making complex constructions more accessible via frequency, greater use of lexical bundles should result in a more readable text. Also, the types of lexical bundles being used in a text are important as well, as they contribute to the organization and coherence of the text. Two research questions are posed here for this portion of the study: (1) What is the relative frequency of lexical bundles in the two texts? and (2) what is the pattern of lexical bundle use in the two texts?

6.2 Methodology

6.2.1 Lexical Bundle Frequency

The frequency of lexical bundles of different lengths was compiled using Unix tools. For the exact process and commands used, see Appendix C. The primary data sources for this portion of the study are frequency lists of lexical bundles of different lengths, examining only bundles that appear more than a minimum number of times. According to Biber, Conrad and Cortes (2004), different minimum frequency cutoffs have been used for analysis, anywhere form 10 occurrences per million words to 40 occurrences per million words (p. 376), and the corpora utilized in such studies are in the range of multiple millions of words. As the present study utilizes a significantly smaller corpus with only two texts, each around 100,000 words, this study will focus on bundles that appear at least five times in either corpus. Bundles of three, four, five, and six words are examined here. This portion of the study will compare both the number of bundles at
each level between the texts, as well as examining the rate of decrease of the number of bundles as length increases.

6.2.2 Functions of Lexical Bundles

It is important first to ascertain the functions of the different bundles. This is done in order to discuss the differences between the texts in the types of bundles used, and the possible effect on readability. There are different means available in the literature for categorizing lexical bundles. The taxonomy to be used here is quoted from Biber, Conrad and Cortes (2004). This taxonomy has the following categories for lexical bundles (pp. 384-387):

1. Stance Expressions: express attitudes or assessments of certainty that frame some other proposition
   Example: if you want to

2. Discourse Organizers: reflect relationships between prior and coming discourse.
   Example: going to talk about

3. Referential Expressions: make direct reference to physical or abstract entities, or to the textual context itself, either to identify the entity or single out some particular attribute of the entity as especially important.
   Example: that’s one of the

As the present study deals with Biblical texts, it is reasonable to assume that the bundles which appear will be somewhat different from those of academic texts or conversation. However, the taxonomies provided here provide a baseline from which to compare the two Bible texts in question. Differences from the taxonomies are, in this case, just as interesting as similarities.
This portion of the study will focus primarily on four-word bundles. Such bundles contain in their structure many of the three word bundles, and they are substantially more frequent than five-word bundles (p. 401). The most frequent bundles from the two texts are compared in the following ways. First, the twenty most frequent four-word bundles were tabulated for each text. Next, the functions of those bundles were ascertained based on the above taxonomy. A further analysis of the most frequent bundles was then conducted, based on sample examples from all of the Gospels in which this bundle appears. These examples were compared with the parallel verses from the other texts, to see if similar bundles appear, and if not, what sorts of constructions are used in its place. Implications for readability will then be discussed later in the chapter.

6.3 Data Analysis

6.3.1 Statistical Comparison of Lexical Bundle Use

As previously stated, the first goal of this portion of the study was to obtain statistical data on the number of lexical bundles of different lengths found in the two texts. Three important trends can be seen in the data, as presented in the table below:

Table 6.1
Table 6.1: Statistical Comparison of Lexical Bundles

<table>
<thead>
<tr>
<th>Number of words</th>
<th>Luther</th>
<th>Hfa</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>types</td>
<td>tokens</td>
</tr>
<tr>
<td>three</td>
<td>1173</td>
<td>10958</td>
</tr>
<tr>
<td>four</td>
<td>321</td>
<td>2858</td>
</tr>
<tr>
<td>five</td>
<td>106</td>
<td>835</td>
</tr>
<tr>
<td>six</td>
<td>35</td>
<td>233</td>
</tr>
</tbody>
</table>

The first trend is clear in both texts: as the length of the bundles increases, the number of bundles found in the texts decreases. This is expected, based on the results of other works on lexical bundles, such as that of Cortes (2004). The second trend is comparative in nature: at all levels Luther contains more types and tokens of lexical bundles than Hfa. Finally the rate at which the number of types and tokens decreases as length increases is greater on average with Hfa than with Luther. In Luther, after each increase in bundle length, the number of remaining types of bundles averages 31.1% of the number of bundles at the previous length, and the number of tokens averages 27.7%. In Hfa, in contrast, the numbers are lower, with an average of 22.9% of types and 20.2% of tokens.

The analysis also shows that lexical bundles account for a much greater proportion of the text in Luther than they do in Hfa. In looking at the most frequent three-word bundles in both, out of a total length of 101,629 words, the top bundle alone accounts for 0.58% of the text. In comparison, the top ranking three-word bundle in Hfa only accounts for 0.096% of the text.

It is also notable when examining the bundles found in both texts that shorter bundles are often part of larger bundles. The table below shows that the highest ranking
bundles from each set for Luther, except for those of length 5, are all part of a bundle of the next longer length.

Table 6.2

<table>
<thead>
<tr>
<th>rank</th>
<th>length</th>
<th>sprach zu ihnen</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>sprach zu ihnen</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>sprach zu ihnen</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>antwortete und sprach zu ihnen</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>jesus antwortete und sprach zu ihnen</td>
</tr>
</tbody>
</table>

This also occurs in Hfa, as can be seen in the next table:

Table 6.3

<table>
<thead>
<tr>
<th>rank</th>
<th>length</th>
<th>vor ihm nieder</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>3</td>
<td>vor ihm nieder</td>
</tr>
<tr>
<td>23</td>
<td>4</td>
<td>sich vor ihm nieder</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>warf sich vor ihm nieder</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>warf sich vor ihm nieder und</td>
</tr>
</tbody>
</table>

Thus, while not always part of larger lexical bundles, it is clear from the above examples that shorter bundles are often foundational in the formation of longer bundles. This has an effect on the length of bundles that have been chosen for functional analysis, as will be seen in the next section.
6.3.2 Lexical Bundles in Context

As stated in the methodology section of this chapter, the following section examines the twenty most frequent four-word bundles from the two texts. As was seen in the examples from the previous section, four-word bundles contain three-word bundles, and the four-word bundles, in turn, are often constituents of five-word bundles, and so on. Four-word bundles have been chosen for this section because they are of more manageable frequency than three-word bundles, and yet still much more prevalent than five-word bundles. Thus, they provide good examples of both high frequency and manageable length, while still exhibiting structural characteristics of interest to be discussed later.

6.3.2.1 Type Trends in Four-word Bundles

The following table shows the top twenty four-word bundles for Luther:

Table 6.4
Table 6.4: Twenty most frequent four-word bundles in Luther

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>92</td>
<td>und</td>
<td>sprach</td>
<td>zu</td>
</tr>
<tr>
<td>2</td>
<td>73</td>
<td>der</td>
<td>war</td>
<td>ein</td>
</tr>
<tr>
<td>3</td>
<td>65</td>
<td>wahrlich</td>
<td>ich</td>
<td>sage</td>
</tr>
<tr>
<td>4</td>
<td>64</td>
<td>und</td>
<td>er</td>
<td>sprach</td>
</tr>
<tr>
<td>5</td>
<td>63</td>
<td>antwortete</td>
<td>und</td>
<td>sprach</td>
</tr>
<tr>
<td>6</td>
<td>62</td>
<td>und</td>
<td>sprach</td>
<td>zu</td>
</tr>
<tr>
<td>7</td>
<td>53</td>
<td>er</td>
<td>sprach</td>
<td>zu</td>
</tr>
<tr>
<td>8</td>
<td>35</td>
<td>jesus</td>
<td>antwortete</td>
<td>und</td>
</tr>
<tr>
<td>9</td>
<td>33</td>
<td>und</td>
<td>es</td>
<td>begab</td>
</tr>
<tr>
<td>10</td>
<td>30</td>
<td>er</td>
<td>aber</td>
<td>sprach</td>
</tr>
<tr>
<td>11</td>
<td>29</td>
<td>aber</td>
<td>sprach</td>
<td>zu</td>
</tr>
<tr>
<td>12</td>
<td>27</td>
<td>der</td>
<td>mich</td>
<td>gesandt</td>
</tr>
<tr>
<td>13</td>
<td>25</td>
<td>wahrlich</td>
<td>wahrlich</td>
<td>ich</td>
</tr>
<tr>
<td>14</td>
<td>24</td>
<td>und</td>
<td>sprachen</td>
<td>zu</td>
</tr>
<tr>
<td>15</td>
<td>23</td>
<td>aber</td>
<td>sprach</td>
<td>zu</td>
</tr>
<tr>
<td>16</td>
<td>21</td>
<td>aber</td>
<td>antwortete</td>
<td>und</td>
</tr>
<tr>
<td>17</td>
<td>20</td>
<td>jesus</td>
<td>aber</td>
<td>sprach</td>
</tr>
<tr>
<td>18</td>
<td>19</td>
<td>sprach</td>
<td>er</td>
<td>zu</td>
</tr>
<tr>
<td>19</td>
<td>18</td>
<td>und</td>
<td>jesus</td>
<td>sprach</td>
</tr>
<tr>
<td>20</td>
<td>18</td>
<td>sprach</td>
<td>zu</td>
<td>ihnen</td>
</tr>
</tbody>
</table>

There are several patterns in the data above. Out of the twenty bundles, seventeen contain one or more verbs that pertain to someone speaking, saying something or answering. Of the remaining three, two begin with the relative pronoun *der* “who”, which refers back to a masculine antecedent, and the other is with the past tense verb *begab* “happened”, which, in this context, means that something occurred.

The following table shows the top twenty four-word bundles for Hfa:

Table 6.5
Table 6.5: Twenty most frequent four-word bundles in Hfa

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>der</td>
<td>mich</td>
<td>gesandt</td>
<td>hat</td>
<td>vor</td>
<td>ihm</td>
<td>nieder</td>
<td>und</td>
<td>jesus</td>
<td>mit</td>
<td>seinen</td>
<td>juengern</td>
<td>in</td>
<td>der</td>
<td>heiligen</td>
<td>schrift</td>
<td>was</td>
<td>ich</td>
<td>euch</td>
<td>sage</td>
</tr>
<tr>
<td>1</td>
<td>22</td>
<td>18</td>
<td>17</td>
<td>16</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>18</td>
<td>17</td>
<td>16</td>
<td>13</td>
<td>13</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>18</td>
<td>17</td>
<td>16</td>
<td>13</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>13</td>
<td>12</td>
<td>11</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>18</td>
<td>16</td>
<td>13</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>18</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>13</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<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>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<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>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<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>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<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>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<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>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<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>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<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>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<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>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<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>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<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>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<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>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<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>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Unlike the Luther text, there is no consistent pattern to the twenty most frequent bundles found in Hfa. Compared to the patterns found in Luther, only three bundles in this group have to do with speaking. Seven of the bundles are either direct or indirect references to people in the text. The remaining ten seem to either describe actions or locations/prepositions, but without much pattern.

### 6.3.2.2 Comparison of Functions of Most Frequent Four-word Bundles

Looking at the five most frequent four-word bundles in Luther and Hfa from the perspective of Biber, Conrad and Cortes’ (2004) taxonomy shows that the two works use bundles for very different purposes. The bundles are structurally different as well.
6.3.2.1 Luther: *und sprach zu ihnen*

The most frequent four-word bundle for the Luther text is the bundle *und sprach zu ihnen* “and spoke to them”. This bundle appears 92 times in the corpus, and is also a part of the second most common five-word bundle in the corpus – *antwortete und sprach zu ihnen* “answered and said to them”. According to the taxonomy presented by Biber, Conrad and Cortes (2004), this bundle fits into the category of discourse organizers.

Some examples of this bundle in context help to illustrate this:

**Matthew 24:4 (Luther):** Jesus aber antwortete *und sprach zu ihnen*: Sehet zu, daß euch nicht jemand verführe.

**Mark 5:39 (Luther):** Und er ging hinein *und sprach zu ihnen*: Was tummelt und weinet ihr? Das Kind ist nicht gestorben, sondern es schläft. Und sie verlachten ihn.

**Luke 3:11 (Luther):** Er antwortete aber *und sprach zu ihnen*: Wer zwei Röcke hat, der gebe dem, der keinen hat; und wer Speise hat, der tue auch also.

**John 2:19 (Luther):** Jesus antwortete *und sprach zu ihnen*: Brechet diesen Tempel, und am dritten Tage will ich ihn aufrichten.

**Acts 27:10 (Luther):** *und sprach zu ihnen*: Liebe Männer, ich sehe, daß die Schifffahrt will mit Leid und großem Schaden ergeben, nicht allein der Last und des Schiffes sondern auch unsers Lebens.

Two things are apparent from these examples. First, this bundle appears throughout the books in the corpus. Second, the function of the bundle is to alert the reader that a message is about to be conveyed, whether by Jesus in the Gospels, or one of the apostles in the book of Acts. It is this signaling aspect that puts the bundle in the category of a
discourse organizer. The reader or the listener, upon receiving this bundle, is alerted to the fact that an important message is about to be conveyed. Hfa, in contrast, does not exhibit such a pattern in the parallel verses. Rather, the role of the bundle is taken typically by a single word, or not at all:

**Matthew 24:4 (Hfa):** Jesus antwortete: »Lasst euch von keinem Menschen täuschen und verführen!


**John 2:19 (Hfa):** Jesus antwortete ihnen: “Zerstört diesen Tempel! In drei Tagen werde ich ihn wieder aufbauen.”

**Acts 27:10 (Hfa):** “Ihr Männer, wenn wir weitersegeln, sehe ich große Gefahren und Schwierigkeiten, und zwar nicht nur für das Schiff und seine Ladung, sondern auch für unser Leben.”

While the use of one word would not increase lexical density for the section, the single word does not provide the familiar signal that the bundles in Luther provide. Commonly used bundles follow a different pattern in Hfa, as will be seen in later sections.

### 6.3.2.2.2 Luther: der war ein Sohn

The bundle *der war ein Sohn* “who was a son” presents us with an interesting pattern. This bundle has a high frequency, appearing 73 times in the corpus, still over three times more than any bundle in Hfa. It differs from the most frequent bundle in Luther, however, in that it has an extremely limited range of verses in which it occurs. In fact, all appearances of this bundle occur in Luther in the third chapter of Luke’s Gospel,
between verses 24 and 38. To say the passage is extremely repetitive would be a great understatement. Here is a sample:

**Luke 3: 24 (Luther):** der war ein Sohn Matthats, der war ein Sohn Levis, der war ein Sohn Melchis, der war ein Sohn Jannas, der war ein Sohn Josephs,

**Luke 3: 25 (Luther):** der war ein Sohn des Mattathias, der war ein Sohn des Amos, der war ein Sohn Nahums, der war ein Sohn Eslis, der war ein Sohn Nangais,

**Luke 3: 26 (Luther):** der war ein Sohn Maaths, der war ein Sohn des Mattathias, der war ein Sohn Simeis, der war ein Sohn Josechs, der war ein Sohn Judas,

**Luke 3: 27 (Luther):** der war ein Sohn Johanans, der war ein Sohn Resas, der war ein Sohn Serubabels, der war ein Sohn Sealthiels, der war ein Sohn Neris,

Hfa, in stark contrast to Luther, does not use a repeated bundle. In fact, the text only contains names and dashes:


**Luke 3: 25 (Hfa):** Mattitja — Amos — Nahum — Hesli — Naggai —


**Luke 3: 27 (Hfa):** Johanan — Resa — Serubbabel — Schealtiel — Neri —

Rather than having any sort of sentences in this section, the reader is left with simply a list of names with no grammatical structure.
6.3.2.2.3 Luther: *wahrlich ich sage euch*

With this bundle, we return to the high use in Luther of discourse organizers in relation to what is about to be said. It is clear from the examples below that each one prefaces a statement of importance:

**Matthew 8:10 (Luther):** Da das Jesus hörte, verwunderte er sich und sprach zu denen, die ihm nachfolgten: *Wahrlich ich sage euch:* Solchen Glauben habe ich in Israel nicht gefunden!

**Mark 9:1 (Luther):** Und er sprach zu ihnen: *Wahrlich ich sage euch:* Es stehen etliche hier, die werden den Tod nicht schmecken, bis daß sie sehen das Reich Gottes mit seiner Kraft kommen.

**Luke 18:17 (Luther):** *Wahrlich ich sage euch:* Wer nicht das Reich Gottes annimmt wie ein Kind, der wird nicht hineinkommen.

**John 6:26 (Luther):** Jesus antwortete ihnen und sprach: *Wahrlich, wahrlich ich sage euch:* Ihr sucht mich nicht darum, daß ihr Zeichen gesehen habt, sondern daß ihr von dem Brot gegessen habt und seid satt geworden.

In the above examples, the reader is clearly made aware that the following statements are of primary importance for the text.

In comparison, the parallel verses from Hfa show no such pattern. Each has a different means of introducing the message to follow:

**Matthew 8:10 (Hfa):** Als Jesus das hörte, wunderte er sich sehr. Er sagte zu den Menschen, die ihm gefolgt waren: »Eins ist sicher: Unter allen Juden in Israel bin ich keinem Menschen mit einem so festen Glauben begegnet.

**Mark 9:1 (Hfa):** Dann sagte Jesus zu seinen Zuhörern: »Das sage ich euch: Einige von euch, die hier stehen, werden nicht sterben, bevor die neue Welt Gottes in ihrer ganzen Kraft sichtbar wird.«

**Luke 18:17 (Hfa):** Hört, was ich euch sage: Wer sich die neue Welt Gottes nicht wie ein Kind schenken lässt, dem bleibt sie verschlossen.
John 6:26 (Hfa): Jesus antwortete ihnen: »Ich weiß, weshalb ihr zu mir kommt: doch nur, weil ihr von mir Brot bekommen habt und satt geworden seid; nicht weil ihr verstanden hättet, was dieses Wunder bedeutet!

While the introductions are different, they do, however, follow a similar pattern in Hfa to that of Luther. Each statement is introduced by a phrase with a colon. The meanings of the phrases are somewhat different, however from Luther’s wahrlich ich sage euch “truly I tell you”. The example of Matthew 8:10 uses Eins ist sicher “one thing is certain”, while the example from Mark uses Das sage ich euch “that I tell you”. The example from Luke gives a directive to the listeners – Hört, was ich euch sage “Hear what I say to you”. The final example from John is a great change – ich weiß, weshalb ihr zu mir kommt “I know why you have come to me”.

6.3.2.2.4 Luther: und er sprach zu

The fourth most frequent bundle in Luther is the first bundle that is not a complete clause. In each case, however, it is clear that this bundle is part of a recurring five-word bundle – und er sprach zu ihnen “and he said to them”. This bundle is the most frequent five-word bundle found in Luther, as well as being a complete clause. The complete bundle occurs 42 times within the text.

Matthew 4:19 (Luther): Und er sprach zu ihnen: Folget mir nach; ich will euch zu Menschenfischern machen

Luke 2:49 (Luther): Und er sprach zu ihnen: Was ist's, daß ihr mich gesucht habt? Wisset ihr nicht, daß ich sein muß in dem, das meines Vaters ist?

John 8:23 (Luther): Und er sprach zu ihnen: Ihr seid von untenher, ich bin von obenher; ihr seid von dieser Welt, ich bin nicht von dieser Welt.

Acts 10:28 (Luther): Und er sprach zu ihnen: Ihr wisst, wie es ein unerlaubtes Ding ist einem jüdischen Mann, sich zu tun oder zu kommen zu einem Fremdling; aber Gott hat mir gezeigt, keinen Menschen gemein oder unrein zu heißen.

Once again, this bundle fills the function of discourse organizer, alerting the reader to the fact that a message is forthcoming in the verse.

The parallel verses from Hfa, similar to the previous set of bundles, often follow the same pattern of phrase and colon, but once again without using a recurring bundle:

Matthew 4:19 (Hfa): Da forderte Jesus sie auf: »Kommt mit mir! Ich will euch zeigen, wie ihr Menschen für Gott gewinnen könnt.«

Mark 3:4 (Hfa): Dann fragte er die Anwesenden: »Soll man am Sabbat Gutes tun oder Böses? Soll man das Leben eines Menschen retten, oder soll man ihn zugrunde gehen lassen?«

Luke 2:49 (Hfa): »Warum habt ihr mich gesucht?«, erwiderte Jesus. »Habt ihr denn nicht gewusst, dass ich im Haus meines Vaters sein muss?«

John 8:23 (Hfa): Dazu sagte ihnen Jesus: »Ihr seid von hier unten; ich komme von oben. Ihr gehört zu dieser Welt; ich gehöre nicht zu dieser Welt.


In this case however, two of the examples deviate from the pattern – the example from Luke, and the example from Acts. In these verses there does not seem to be a similar
phrase, and the structure is rather different, beginning with a quote instead of first pre-facing it with a phrase and colon.

6.3.2.2.5 Luther: *antwortete und sprach zu*

The final bundle to be examined from Luther is one that actually spans two clauses. While the examples below show different surrounding contexts, this bundle is a component of the second most frequent five-word bundle in Luther: *antwortete und sprach zu ihnen* “answered and spoke to them”. Despite the differences between the examples, it is clear that they are still functioning as discourse organizers, once again introducing important statements:

**Matthew 13:37 (Luther):** Er *antwortete und sprach zu ihnen:* Des Menschen Sohn ist's, der da Guten Samen sät.

**Mark 7:6 (Luther):** Er aber *antwortete und sprach zu ihnen:* Wohl fein hat von euch Heuchlern Jesaja geweissagt, wie geschrieben steht: "Dies Volk ehrt mich mit den Lippen, aber ihr Herz ist ferne von mir.

**Luke 4:12 (Luther):** Jesus *antwortete und sprach zu ihm:* Es ist gesagt: "Du sollst Gott, deinen HERRN, nicht versuchen."

**John 2:19 (Luther):** Jesus *antwortete und sprach zu ihnen:* Brechet diesen Tempel, und am dritten Tage will ich ihn aufrichten.

The parallel verses from Hfa show the first evidence of a pattern, one which typically is a shortened version of the bundle found in Luther. Rather than saying *antwortete und sprach zu* “answered and spoke to”, the Hfa examples often reduce this to one word – *antwortete* “answered”:

**Matthew 13:17 (Hfa):** Jesus *antwortete:* »Der Menschensohn selbst ist der Bauer, der die gute Saat aussät."
Mark 7:6 (Hfa): Jesus antwortete: »Wie Recht hat Jesaja, wenn er von euch Heuchlern schreibt: ›Diese Leute ehren Gott mit den Lippen, aber mit dem Herzen sind sie nicht dabei.

Luke 4:12 (Hfa): Aber Jesus wies ihn auch diesmal zurück: »Es steht aber auch in der Schrift: ›Du sollst Gott, deinen Herrn, nicht herausfordern!«

John 2:19 (Hfa): Jesus antwortete ihnen: »Zerstört diesen Tempel! In drei Tagen werde ich ihn wieder aufbauen.«

6.3.2.2.6 Hfa: der mich gesandt hat

The most frequent type of four-word bundle lexical bundle in Hfa is much different from that of Luther. Rather than being a discourse marker, der mich gesandt hat “who has sent me” has a referential function. It is not as widely distributed in the text as und sprach zu ihnen “and spoke to them” is in Luther: it only appears in three of the four Gospel, with the highest concentration in the book of John, and it does not appear at all in the book of Acts. The examples below make it clear how der mich gesandt hat functions as a referential bundle:

Matthew 10:40 (Hfa): Wer euch aufnimmt, der nimmt mich auf, und wer mich aufnimmt, der nimmt Gott auf, der mich gesandt hat.


John 6:38 (Hfa): Denn ich bin nicht vom Himmel gekommen, um zu tun, was ich will, sondern um den Willen des Vaters zu erfüllen, der mich gesandt hat.

In each of the above examples, der mich gesandt hat functions as an anaphoric reference to a proper noun referring to God, which is highlighted in the examples along with the bundles themselves. As above, I have compared these verses with
the corresponding verses in Luther. In each example below, we see that the same bundle appears in each example:

**Matthew 10:40 (Luther):** Wer euch aufnimmt, der nimmt mich auf; und wer mich aufnimmt, der nimmt **den** auf, **der mich gesandt hat.**

**Luke 10:16 (Luther):** Wer euch hört, der hört mich; und wer euch verachtet, der verachtet mich; wer aber mich verachtet, der verachtet **den**, **der mich gesandt hat.**

**John 6:38 (Luther):** Denn ich bin vom Himmel gekommen, nicht daß ich meinen Willen tue, sondern den Willen **des**, **der mich gesandt hat.**

While the same bundle appears, there are some key differences between the above examples from Luther and those from Hfa. Where the antecedent is clearly stated in the examples from Hfa, the corresponding Luther examples refer back to pronouns, whose antecedents do not appear within the verses. In this sense, Hfa has made it clearer for the reader, especially one who is unfamiliar with the Bible, to whom the bundle is referring. This is by no means the case in all examples of this bundle for Hfa, but it is interesting that the two works correspond in this way.

6.3.2.2.7 Hfa: *vor ihm nieder und*

The second most frequent bundle from Hfa, *vor ihm nieder und* “before him down and” is a clear example of a referential bundle that refers to a location.

**Matthew 2:11 (Hfa):** Sie gingen in das Haus, wo sie das Kind mit seiner Mutter Maria fanden, knieten **vor ihm nieder und** ehrten es wie einen König. Dann packten sie ihre Schätze aus und beschenkten das Kind mit Gold, Weihrauch und Myrrhe.

**Mark 5:33 (Hfa):** Die Frau war erschrocken und zitterte am ganzen Leib, denn sie wusste ja, was an ihr geschehen war. Sie fiel **vor ihm nieder und** sagte ihm alles.

John 11:32 (Hfa): Kaum hatte er Jesus gesehen, fing er an zu schreien. Er warf sich vor ihm nieder und rief laut: »Was willst du von mir, Jesus, du Sohn Gottes, des Höchsten? Ich flehe dich an, quäle mich nicht!«

Two of the four parallel verse examples from Luther contain the same bundle as Hfa:

Matthew 2:11 (Luther): und gingen in das Haus und fanden das Kindlein mit Maria, seiner Mutter, und fielen nieder und beteten es an und taten ihre Schätze auf und schenkten ihm Gold, Weihrauch und Myrrhe.

Mark 5:33 (Luther): Das Weib aber fürchtete sich und zitterte (denn sie wußte, was an ihr geschehen war), kam und fiel vor ihm nieder und sagte die ganze Wahrheit.


John 11:32 (Luther): Als nun Maria kam, da Jesum war, und sah ihn, fiel sie zu seinen Füßen und sprach zu ihm: HERR, wärest du hier gewesen, mein Bruder wäre nicht gestorben!

In the frequency rankings, this bundle appears in 108th place in Luther’s translation. It does, however appear seven times in the text, such that it is still part of the analysis for this chapter. In comparison, only the 19th through the 33rd most frequent four-word bundles appear 7 times in Hfa.

6.3.2.2.8 Hfa: Jesus mit seinen Jüngern

This next bundle continues the pattern of Hfa using primarily referential bundles. This time, the bundle is framed with two nouns – Jesus mit seinen Jüngern “Jesus with
his disciples”. This bundle does not appear as part of any larger bundles. In each case we simply have a reference to who Jesus was with in a given situation.


In the parallel verses in Luther, we do not see an exact bundle repeated here. Typically, if both Jesus and the disciples are mentioned, they are conjoined with und “and”, rather than the preposition mit “with”. Other times, as in the example from Luke, Jesus is only referred to with the pronoun er “he”:

Matthew 9:10 (Luther): Und es begab sich, da er zu Tische saß im Hause, siehe, da kamen viele Zöllner und Sünder und saßen zu Tische mit Jesu und seinen Jüngern.

Mark 2:15 (Luther): Und es begab sich, da er zu Tische saß in seinem Hause, setzten sich viele Zöllner und Sünder zu Tische mit Jesu und seinen Jüngern; denn ihrer waren viele, die ihm nachfolgten.


John 3:22 (Luther): Darnach kam Jesus und seine Jünger in das jüdische Land und hatte daselbst sein Wesen mit ihnen und taufte.
6.3.2.2.9 Hfa: *in der heiligen Schrift*

This bundle in Hfa fills another type of reference, being that it refers to a text as well as to a location. This bundle is also part of a larger five-word bundle – *in der Heiligen Schrift vorausgesagt* “stated in the Holy Text”.

**Matthew 4:6 (Hfa):** »Spring hinunter!«, forderte er Jesus auf. »Du bist doch Gottes Sohn! Und *in der Heiligen Schrift* steht: ›Gott wird seine Engel schicken. Sie werden dich auf Händen tragen, und du wirst dich nicht einmal an einem Stein verletzen!‹ «

**Mark 11:17 (Hfa):** »Ihr wisst doch, was Gott *in der Heiligen Schrift* sagte, rief Jesus der Menschenmenge zu: »›Mein Haus soll für alle Völker ein Ort des Gebets sein‹, ihr aber habt eine Räuberhöhle daraus gemacht.«

**Luke 4:4 (Hfa):** Aber Jesus wehrte ab: »Nein, denn es steht *in der Heiligen Schrift*: ›Der Mensch lebt nicht allein von Brot, sondern von allem, was Gott ihm zusagt!‹ «

**John 2:17 (Hfa):** Seine Jünger aber mussten an das Wort *in der Heiligen Schrift* denken: »Der Eifer für deinen Tempel wird mich vernichten!«

**Acts 13:29 (Hfa):** Als sie alles getan hatten, was *in der Heiligen Schrift* vorausgesagt ist, nahmen sie ihn vom Kreuz herunter und legten ihn in ein Grab.

In the examples from Luther, there is a common participle that stands in place of the bundle found in Hfa: *geschrieben* “written”, or often in combination with the verb *stehen* “to stand”, meaning “it is written”. In no case in the examples does the Luther version say explicitly where the statement is written:

**Matthew 4:6 (Luther):** und sprach zu ihm: Bist du Gottes Sohn, so laß dich hinab; denn *es steht geschrieben*: Er wird seinen Engeln über dir Befehl tun, und sie werden dich auf Händen tragen, auf daß du deinen Fuß nicht an einen Stein stoßest.

**Mark 11:17 (Luther):** Und er lehrte und sprach zu ihnen: *Steht nicht geschrieben:* "Mein Haus soll heißen ein Bethaus allen Völkern"? Ihr aber habt eine Mördergrube daraus gemacht.
Luke 4:4 (Luther): Und Jesus antwortete und sprach zu ihm: **Es steht geschrieben:** "Der Mensch lebt nicht allein vom Brot, sondern von einem jeglichen Wort Gottes."

John 2:17 (Luther): Seine Jünger aber gedachten daran, daß **geschrieben steht:** Der Eifer um dein Haus hat mich gefressen.

Acts 13:29 (Luther): Und als sie alles vollendet hatten, was von ihm **geschrieben** ist, nahmen sie ihn von dem Holz und legten ihn in ein Grab.

6.3.2.2.10 Hfa: **was ich euch sage**

The final bundle to be examined from Hfa – **was ich euch sage** “what I say to you” – is the only one in the group to show the sort of discourse organizational role that we have seen in many of the Luther bundles. In most cases, however, it does so differently from Luther. Where discourse organizers in Luther typically come early in a verse, most occurrences of this bundle in Hfa are at the end, as can be seen in the following examples:

**Matthew 11:15 (Hfa):** Hört genau auf das, **was ich euch sage.**

**Mark 10:15 (Hfa):** Hört, **was ich euch sage:** Wer sich die neue Welt Gottes nicht wie ein Kind schenken lässt, dem bleibt sie verschlossen.

**Luke 6:46 (Hfa):** Warum nennt ihr mich dauernd ›Herr!‹, wenn ihr doch nicht tut, **was ich euch sage?**

**John 5:47 (Hfa):** Wenn ihr aber nicht einmal glaubt, was er geschrieben hat, wie könnt ihr dann glauben, **was ich euch sage?**

Only in the example from Mark do we see the bundle in a position to introduce a statement. This suggests, that this bundle serves two different functions – discourse organizing in the Mark example, and text reference in the other, referring to other statements of the speaker.
Parallel verses in Luther only use the bundle once, and it appears as a four- or five-word bundle less than five times in the entire text. It can be noted, however, that *wahrlich ich sage euch* “truly I tell you” does appear in the example from Mark.

**Matthew 11:15 (Luther):** Wer Ohren hat, zu hören, der höre!

**Mark 10:15 (Luther):** Wahrlich ich sage euch: Wer das Reich Gottes nicht empfängt wie ein Kindlein, der wird nicht hineinkommen.

**Luke 6:46 (Luther):** Was heißt ihr mich aber HERR, HERR, und tut nicht, *was ich euch sage*?

**John 5:47 (Luther):** So ihr aber seinen Schriften nicht glaubt, wie werdet ihr meinen Worten glauben?

### 6.4 Discussion

In this portion of the study, I examined two aspects of lexical bundles in the two texts, one which is quantifiable, and the other more subjective. Both aspects, however, have a potential impact on the readability of the texts. The following discussion shall address the quantitative and qualitative analyses in turn, followed by a discussion of how the two interact, as well as some possible implications for readability.

#### 6.4.1 Comparative Frequency and Variety of Lexical Bundles

Based on the data above, the Luther text utilizes many more lexical bundles than Hfa. Thus, Luther has a potential readability advantage over Hfa, as in each category of bundles there are several times more present in the older text. Also, high frequency bundles at each level appear often as constituents in correspondingly high frequency
bundles of the next longer set. While this also occurs in Hfa, the example shown from Luther above in table 6.2 shows that the first most frequent three-word bundle was a part of all most frequent bundles except for the five-word bundle, and even then it was in second place. Thus, the readers of Luther are being shown the same bundles in a frequent and consistent manner throughout the text to a much greater degree than they are in Hfa. Indeed, Biber and Conrad (1999), place great importance on the frequency of lexical bundles, calling them “basic building blocks for both spoken and written discourse” (p. 188), due to their frequency. It is clear that the text of Luther is founded upon far more of these blocks than Hfa. The readers of Luther thus have many more highly frequent markers guiding them through the text.

It is also important to understand what role lexical bundles play in making a text easier to read. We have already seen one role in the statements from Biber and Conrad (1999), in that it eases the reading and understanding of complex structures. Each bundle of high frequency serves as a marker of some sort in the text, but the purpose of such markers should also be considered, in order to understand exactly how the bundles are aiding the reader. This will be discussed in the next section.

6.4.2 Comparative Types and Functions of Lexical Bundles

In examining the most frequent four-word bundles from the two texts, very different patterns emerge. Both, however, have benefits for the readability of the text. As we saw in sections 6.3.2.2.1-5, the top four-word bundles for Luther were mainly discourse organizational bundles. I categorized these bundles this way due to the fact
that they alert the reader to an upcoming message from one of the key speakers in the corpus, typically, as seen from the complete verses, as a response to a question or a situation. It is striking just how similar most of the top five bundles are:

\begin{verbatim}
und sprach zu ihnen
der war ein Sohn
wahrlich ich sage euch
und er sprach zu
antwortete und sprach zu
\end{verbatim}

With the exception of \textit{der war ein Sohn} “he was a son”, each bundle contains a verb that indicates speaking or telling, and these are in bold print above. Great emphasis is placed in Luther on what is to be said, and in announcing it\textsuperscript{2}.

An interesting characteristic of the \textit{und sprach zu ihnen} and \textit{wahrlich ich sage euch}, as compared to other bundles observed both in the Luther and Hfa texts and by researchers such and Biber (1999), is that they are complete clausal units as well as lexical bundles. Typically, bundles are not such complete units. Rather, they more often span clausal boundaries and contain only fragments of complete clauses. According to Biber and Conrad (2004), only 15\% of all lexical bundles in conversation and 5\% of lexical bundles in academic prose can be considered as complete clausal units (p. 377). Therefore, these most frequent bundles in Luther are something rather special. They give the reader the opportunity to reduce entire clauses to one single unit.

Although the remaining two bundles discussed above do not constitute complete clauses, they fulfill extremely similar roles as discourse organizers, both on their own and

\textsuperscript{2} The second most frequent bundle – \textit{der war ein Sohn} – is anomalous both for its referential nature and its highly restricted area of appearance.
as components of larger bundles, which do comprise complete clauses. This should have two particular effects on readability. First, by reducing the clause to one unit, it reduces the clausal complexity of a given sequence, making the passage easier to read. Each time one of these bundles appears, rather than being seen as a new clause by the reader, it can be seen a single entity, after which the reader can move quickly on to the more important information to follow. This connects to the importance of lexical bundle function, and the second aid to readability. As a discourse organizer, this unit functions as a clear marking that important information will follow the bundle in the text.

In contrast to Luther, Hfa only uses single verbs in the parallel verses shown above. This can be seen clearly in the following comparison:

**Matthew 11:4 (Hfa):** 4 Jesus antwortete: »Geht zu Johannes zurück und erzählt ihm, was ihr hört und seht:

**Matthew 11:4 (Luther):** 4 Jesus antwortete und sprach zu ihnen: Gehet hin und saget Johannes wieder, was ihr sehet und höret:

While this should not detract from the readability of the text, I believe it also does not add to its readability, as the repeated bundles found in Luther could also be considered as single units, processed as a whole and not as individual words. The lexical bundles seen in Luther, as they are longer orthographically, should provide a stronger, more salient indication that important information is coming. This also speaks to the more oral nature of the Luther translation. According to Ong (1982), orally transmitted traditions and knowledge has a tendency towards being “redundant or ‘copious’”. This can indeed be seen in the Luther example, where two verbs antwortete “answered” and sprach “spoke” are utilized, when both mean that Jesus verbally expressed something (p. 39).
Redundancy is also an aspect of “oral expression before a large audience” (p. 40), which is also a venue in which the Bible is often experienced.

This is not to say that Hfa does not use lexical bundles in a way that improves readability. In the examples above in section 6.3.2.2.2, we see that Hfa’s use of referential bundles is actually clearer than that of Luther. It is helpful here to have a direct comparison from the examples:

**Matthew 10:40 (Hfa):** Wer euch aufnimmt, der nimmt mich auf, und wer mich aufnimmt, der nimmt Gott auf, der mich gesandt hat.

**Matthew 10:40 (Luther):** Wer euch aufnimmt, der nimmt mich auf; und wer mich aufnimmt, der nimmt den auf, der mich gesandt hat.

As noted above, both verses use the exact same bundle: *der mich gesandt hat.* The difference is in the proximity of a proper noun antecedent for that bundle. In Hfa, that antecedent is found within the same verse, in the previous clause – the noun *Gott* “God”. In Luther, however, the most immediate antecedent for the bundle is the relative pronoun *den.* A proper noun antecedent is not found within the verse. Surrounding verses in Luther are topically different as well, and do not contain the antecedent. Thus, by using the proper noun *Gott* “God”, Hfa strengthens the referential nature of the bundle by making explicit an antecedent that is only implied in the Luther text. On further examination of the examples found in 7.3.2.2.2, it is clear that this contrast is present in each of the comparisons of parallel verses.

Another example of where a reference is clearer in Hfa than in Luther is with the example *in der Heiligen Schrift* “in the Holy Scripture”. In these examples in Hfa, it is clearly referenced where the source for the speaker’s statement can be found. A direct comparison of two parallel verses shows that in Luther, this reference is not so clear:
Mark 11:17 (Hfa): »Ihr wisst doch, was Gott in der Heiligen Schrift sagt«, rief Jesus der Menschenmenge zu: »›Mein Haus soll für alle Völker ein Ort des Gebets sein‹, ihr aber habt eine Räuberhöhle daraus gemacht.«

Mark 11:17 (Luther): Und er lehrte und sprach zu ihnen: Steht nicht geschrieben: "Mein Haus soll heißen ein Bethaus allen Völkern"? Ihr aber habt eine Mördergrube daraus gemacht.

The Luther text says only that the reference of the speaker steht geschrieben “stands written”, but gives no location or direct reference to a specific source. Once again, the reader is forced to resort to outside knowledge in order to infer where something is written, rather than having it explained directly in the text.

In order to consider the implications for readability from these patterns, one must take into account the experience of the reader. An experienced reader and interpreter of the Bible would probably be aware that the relative pronouns found in the Luther text refer back to some incarnation of God. For a novice reader of the book, however, such unclear anaphora could prove confusing, making the reading of the text more difficult. In this sense, Hfa’s use of a common lexical bundle in tandem with close anaphoric reference should increase the readability of the text for novice and experienced reader alike. It is a similar situation with the statement of in der Heiligen Schrift “in the Holy Scripture”. An experienced reader could interpret from Luther that steht geschrieben “stands written” implies that the reference is written in the Holy Text. A novice reader, however, would have no such point of reference, and be left to wonder where the message was written. Hfa makes this clear, further lending to the coherence of the passage.

It is clear from these two examples and the others cited above from Hfa that great importance is placed in the use of lexical bundles in the function of reference, especially
reference to location. The repetition of bundles like *vor ihm nieder und* “down before him and” and *Jesus mit seinen Jüngern* “Jesus with his disciples” indicates that emphasis is placed heavily on the location of others in reference to Jesus. Where the Luther text would use discourse organizing bundles to indicate that important messages were forthcoming, Hfa seems to use them as important referential messages in and of themselves. This could serve to increase the readability of the text by acting as a marker of what the reader should see as key ideas in the text.

### 6.4.3 Lexical Bundles and Readability

As discussed in the introduction to this chapter, lexical bundles are fundamental building blocks of both written and spoken language. In this portion of the study, I examined lexical bundles from two distinct angles – from a purely quantitative angle, and from a functional angle. Both have important consequences in regards to the readability of the texts in question. In looking at the two research questions stated in the introduction to this chapter, the two texts do indeed have different levels of bundle frequency and differing patterns of functional use of lexical bundles.

Readability is aided by more frequent use of lexical bundles because higher frequency of a greater number of bundles potentially means that more of the bundles are familiar to the reader and thus function as individual units, rather than separate words. From this perspective, Luther has the clear advantage over Hfa, in that Luther contains many more examples of lexical bundles than its modern counterpart. This is consistent for all of the bundles examined, regardless of length. Luther’s use of lexical bundles also
adds to the cohesion of the text by informing the reader in a very clear and consistent manner that key statements are following the bundles used.

Despite its less frequent use of lexical bundles, Hfa does show examples of where the lexical bundles that it does use are used sometimes more effectively than those in Luther. By providing a proper noun antecedent in the same verse, Hfa provides clarity to the reader, where Luther leaves the antecedent somewhat ambiguous by referencing another pronoun. According to Halliday and Matthiessen (2004), reference “creates cohesion by creating links between elements” (p. 534). The examples of referential bundles from Luther require the reader to have knowledge outside of the immediate text – that is, who is represented by the demonstrative pronouns present in the preceding clauses. The reader of Hfa requires no such external knowledge, since there is a definite antecedent present. In general, the reader should be aided by the referential function of bundles in Hfa – by knowing where important references are to be found, and by knowing what sorts of locations in relation to Jesus are important. This is an important point, as it indicates not only possible differences in biblical literacy but also facilitation via the use of short passages.

6.5 Conclusion

Several conclusions can be made from the results of the analyses. First and foremost, the considerably higher usage of lexical bundles, by this metric, indicates that the Luther text was more readable for his contemporaries than Hfa is for modern readers. Luther’s use of lexical bundles is far more frequent, far more formulaic, and far more
purposeful. By formulaic and purposeful, I point to the pattern that Luther’s most frequent lexical bundles fall mainly into the category of discourse organizers. The reader is given large, repetitive bundles that indicate that important messages are coming in the text. Examples, such as wahrlich ich sage euch “truly I tell you” provide the reader with a familiar point of direction. The condensation of such frequent bundles into single units provided the 16th century reader with a text that was both less dense and easier to follow. Hfa does not consistently use such bundles.

The results of the quantitative and functional analyses of the two texts raise important questions for further study. The examination of individual lexical bundles indicates that greater frequency alone is not necessarily indicative of an increase in readability. In the case of the referential bundles in Luther, the lack of a clear antecedent would seem to work against readability despite the frequency of the lexical bundle. Further analysis of individual reading times and an analysis of readers’ impressions of the helpfulness of lexical bundles would go far in explaining this hypothesis.
Chapter 7

Conclusion

7.1 Introduction

Dewey (2004) posed the question that is at the core of this project on German Bibles: “is it readable?” (p. 33). As discussed in chapter 1, translators of both versions sought to produce a Bible that was readable and accessible to their contemporaries. Chapter 2 presents an overview of different approaches to measuring the readability of texts. Chapters 3 through 6 applied statistical measures of aspects of readability to the selected texts from Luther and Hfa in order to determine which version was more readable for its contemporaries. A summary and discussion of the findings concludes the dissertation.

7.2 Summary of Findings

7.2.1 Lexical Density

Following the means of calculating lexical density developed by Ure (1971) and used by Stubbs (2000), lexical density was measured as the proportion of lexical words to total words in the two corpora. This was done at four levels: comparing parallel verses, chapters, books, and the complete corpora. Each level’s lexical density was also calculated in two ways, the first including all types of verbs regardless of function, and
the second excluding modal and auxiliary verbs. At each level and with both means of calculation, statistically significant differences were found, showing that the Luther text has consistently lower lexical density than Hfa.

Sections 3.3.2.1 and 3.3.2.2 continued by taking into account the concentration of specific types of lexical words within the total number of content words, specifically looking at nouns and verbs. The concentrations were given as the percentage of lexical words that were nouns and the percentage that were verbs. Once again, the calculations were conducted in two ways, both including modals and auxiliaries in the verb count and total lexical word count, and excluding those verbs. These data were then analyzed in comparison to the lexical density data in order to ascertain whether or not there was a correlation between noun concentration and lexical density and verb concentration and lexical density. No significant correlation was found between noun concentration and lexical density in either text.

Strong negative correlations, however, were found between verb concentration and lexical density with both means of calculation. This indicates that the more verbs there are in a text, the lower its lexical density. Statistical comparisons of verb concentration with both measures of lexical density indicated significant differences between the two texts, with Luther containing the higher concentration of verbs. Due to the negative correlation between verb concentration and lexical density, this is a further indication of the lower lexical density of the Luther text.
7.2.2 Clausal Complexity

The study of specific parts of speech continued in chapter 4, looking at the frequency and concentrations of different types of clausal markers, focusing on coordinating and subordinating conjunctions. An interesting aspect of these data was the overwhelming frequency in both texts of the conjunction *und* “and”, but most especially in the Luther text. Conjunctions were examined at three levels: chapter, book, and the complete corpus. Once again, calculations were done in two ways, both including *und* and excluding *und*, due to its high frequency and ability to appear in places other than clausal boundaries. Two means of statistical analysis were employed here. First, the corpora were examined for significant differences in the frequencies of conjunctions. Second, conjunctions were looked at as a percentage of total words in the text and this value was compared via correlation statistics with corresponding values for lexical density at those levels.

With the inclusion of *und*, significant differences were found at all levels for the concentration of coordinating conjunctions between Luther and Hfa, with Luther having a consistently higher frequency of conjunctions compared to Hfa. When *und* “and” was excluded from the calculation, significant differences were found at the book level, but not at the chapter level. For subordinating conjunctions, significant differences were found at both the book and chapter levels, with Luther having the higher frequency of conjunctions in both. Statistical analysis in this chapter examined the correlation between coordinating conjunctions, both with and without *und*, and with lexical density calculated with and without modals and auxiliaries. A significant negative correlation
was found, once again indicating, as with the verbs in chapter 3, that a higher frequency of conjunctions indicates lower lexical density. Significant negative correlations were also found when looking at concentrations of subordinating conjunctions with both calculations of lexical density.

### 7.2.3 Nominalization

The examination of nominalizations used in the two corpora took place in two stages. Both were focused on nominalizations marked with specific suffixes. The first part looked at the frequencies nominalizations containing one of four different suffixes: -ung, -heit, -keit, and –nis. The ten most frequent word types from these categories were compared, looking specifically at types which overlapped between the two corpora. Log-likelihood analyses of these overlapping types showed that most did not have significantly different frequencies, except for one type in the –nis category. When taken as groups however, there were significant differences between all suffix groups except –keit, with Hfa having consistently higher frequencies of nominalizations compared to Luther.

This section continued by analyzing appearances of some of the most frequent –ung nominalizations in context. These nominalizations stand out from the others in two ways: first, they are by far the most frequent in both corpora, and second, they are the only group examined that was derived from verbs, while the other nominalizations were derived from adjectives. This is an important distinction, as the nominalization of processes is the key issue in work such as that of Halliday (2004). The goal of this
section was to examine the context in which these nominalizations appear, looking for agnate or other forms of clarification. There was no consistent pattern in which one version provided more clarification than another, when examining nominalizations that had high frequencies of occurrence in both texts. With nominalizations that did not appear in both texts, Hfa did show more clarification, or at least the replacement of some nominalizations with others that are more frequent in the current German lexicon.

7.2.4 Lexical Bundles

In Chapter 6, lexical bundles were the focus of analysis. Lexical bundles are frequently recurring strings of words. In the case of this corpus, strings that occurred at least five times were examined. The study of lexical bundles took place in two stages. The first examined the frequencies of lexical bundles of different lengths, ranging from three to six words. Lexical bundles were found to be more frequent in Luther, whether measured by types or tokens of these bundles.

The second stage of this chapter looked specifically at the most frequently occurring four-word bundles, in order to determine how these bundles functioned in context, and how they aided the reader. In Luther, the most frequent bundles were overwhelmingly used as discourse organizational markers. Such bundles indicate to the reader that an important message is coming in the text. Such bundles were rarely found in Hfa, which showed more of a tendency towards referential bundles. In comparing the use of such referential bundles, however, it appeared that Hfa referred back to nouns more often, whereas references in Luther often referred most directly to pronouns.
Consequently, Hfa’s use of referential bundles was clearer and more helpful for the reader. Thus, while Luther’s bundles have the benefit of frequency, Hfa often shows more effective use of such bundles, when they are present.

7.3 Implications of Findings for Readability

Having examined the results of each section, some conclusions can be made in regard to the overall research questions. For the first question, it is clear that there are differences between the two texts, indicating that one text was more readable for its contemporaries than the other.

For lexical density, significantly lower lexical density indicates better readability. This was the case for Luther at all levels, both with the inclusion of modals and auxiliaries in the set of lexical words and when they were excluded. There is an inverse correlation between lexical density and verb concentration. Thus, a high concentration of verbs correlates with low overall lexical density. Luther’s translation had a consistently higher concentration of verbs, both with and without modals and auxiliaries included in the verb count. There is a historical tendency towards greater use of auxiliary and modal verbs in German. Therefore, it is surprising that Luther shows a higher concentration of verbs even when auxiliaries and modals are counted.

Clausal complexity is associated with the differences between spoken and written language. Halliday and Matthiessen (2004) argue that written language is more lexically dense, and spoken language is more grammatically complex. Greater clausal complexity is an indication of greater grammatical intricacy. Greater clausal complexity in a
translation, then, more closely approximates spoken language, and is therefore less lexically dense. As we saw from the previous section, lower lexical density indicates better readability. With und “and” included in the calculations, Luther has overwhelmingly and significantly more coordinating conjunctions than Hfa at all levels. This is less clear when und “and” is excluded, although at the book level, Luther does maintain significantly higher concentrations of conjunctions. Luther also contains significantly more subordinating conjunctions at all levels. Concentrations of both coordinating and subordinating conjunctions show significant negative correlation with lexical density, once again leading to the conclusion that Luther, by containing more conjunctions, is less lexically dense and therefore more readable for its contemporaries.

The issue is less clear when examining nominalization. From the standpoint of frequency, Luther has significantly fewer occurrences of most of the suffix-types examined. As nominalizations are typically classed as difficult word types, lower frequency should equal better readability. When looking at the use of nominalizations in context, however, it was seen that neither text was consistent in clarifying such nominalizations, and in the overlapping nominalizations examined, the differences in frequency were not significant. Therefore, Luther did not aid readability by consistently clarifying such nominalizations via the use of agnate forms. Hfa, in turn, did not improve upon this Luther, in the case of overlapping nominalizations. In the cases of non-overlapping nominalizations, Hfa did show some improvement in readability by either using agnate forms in place of the nominalizations found in Luther, or by using nominalizations with higher frequency in the modern German lexicon. Therefore, Luther
has the advantage in readability because of a lower frequency of nominalization, but actual nominalization usage is problematic for both.

Finally, with lexical bundles, we see overwhelming statistical support for the better readability of Luther. More highly frequent lexical bundles indicate more strings of words that the reader is familiar with and can process as single units. Luther showed consistently higher frequencies of all lengths of lexical bundles examined here, as well as more consistent use of lexical bundles at the four-word length. Hfa, however, did show more helpful use of referential bundles. Once again, as with nominalizations, frequency of lexical bundles indicates that Luther’s translation is the more readable text.

The statistical and analytical measures performed in this dissertation indicate that the Luther text was more readable for its contemporaries than Hfa for the people of the late 20th and early 21st centuries. The question now becomes, what drove these changes between Luther and Hfa? Were the differences prompted by changes in the German language, or by changes in the readership of the Bible and the ways in which the Bible is received? While we do see some changes in the language between the two texts, I believe that the differences are more strongly due to differences between how the Bible is received now and how it was received in the 16th century. As was discussed in chapter 1, there have been great changes in the literacy levels of Germans since the 16th century, rising from only 4% literacy to 99%. This is coupled with a currently low percentage of people who claim to be part of the mainline evangelical church in Germany. Thus, while more people can read the Bible on their own, only a small percentage of them choose to receive it in public today, compared to Luther’s time, in which the majority of the people receiving the Bible could only do so aurally. The tendencies that help Luther to have
lower lexical density are often directly connected to the necessarily oral nature of that translation.

Three aspects of oral tradition in general from Ong (1982) are seen as important here: 1) the additive nature of oral language, 2) the aggregative nature of oral language, and 3) the redundant nature of oral language. The first, the additive nature of oral language, is clearly seen in the negative correlation between the concentration of coordinating conjunctions and lexical density. The overwhelming use by Luther of *und* “and” alone is evidence of this additive tendency. Both the aggregative and redundant natures of oral language are clearly seen in the data on lexical bundles in Luther. These bundles are at once formulaic, frequent, and redundantly phrased, such as with *er antwortete und sprach* “he answered and spoke”. The use of fewer nominalizations in Luther is further evidence of orality, as this is associated more with spoken language. Lower lexical density, and therefore greater readability, then, is tied to the oral nature of the Luther translation.

7.4 Avenues for Further study

Research into the comparative readabilities of these two texts is by no means complete. Each chapter has areas in which more in-depth and nuanced research could be done. The corpus itself could be very easily expanded in order to make generalizations based on more data. The current corpus contained only five books out of eighty found in Luther, including apocrypha. While the books included in this work were narrative in nature, a number of other genre types can be found within the Bible, including epistles.
and poetry. Further study into readability in the Bible would have to be adapted to these
differences.

From the specific chapters, more work could especially be done with lexical
density and clausal complexity. While this study looked at lexical density as a proportion
of lexical words to total words, Halliday and others take a more nuanced approach to the
construct. Halliday and Matthiessen measure lexical by dividing “the number of lexical
items by the number of ranking clauses” (Halliday and Matthiessen, 2004, p. 655). In
terms of corpus analysis, this would require more nuanced tagging of clausal boundaries
and clause types, thus combining the work of chapters 3 and 4 into a more coherent
whole.

Another area for more analysis would be types of nominalization. This would
again require more nuanced tagging. The POS tag set used for this study included only
two categories for nouns: regular nouns and proper nouns. A further study of
nominalization would require tagging that differentiated between different types of
nominalizations and nouns, thus making accessible nominalizations with a wider array of
suffixes, as well as those with no suffixes at all.

Finally, these methods can be applied to other Bible versions, both in German and
in English. Similar studies of historically disparate versions could be done, as well as
studies of contemporary Bibles as seen in the work of Resnik (1999).

This research also sets a foundation for more critical analysis of the Bibles
examined here. As seen in chapter 1, much of the criticism leveled at Hfa is based on
linguistic data, albeit somewhat limited. According to the publishers of Hfa, much of the
criticism against their translation is based on limited data and limited knowledge of the
translational methods used to create Hfa (http://www.hoffnungfueralle.com, last accessed on 30-03-2007). They welcome, however, critique and discussion that looks not only at the results of the translation of Hfa, but at the classical translations that critics use as the basis for their arguments (ibid.). Critique based on a systematic study of readability should help to move the conversation beyond a comparison of the established norms for Bible translations and the problems of a new Bible translation. Instead, such a study would continue from the foundation of the work done in this dissertation. Rather than having arguments based on few or individual instances of change, such a study could base critique on statistically significant issues found in the texts, such as those found in this work.

Clearly, there is great potential for further study based on the work done here. This dissertation is a first step. The results of the four analyses – lexical density, clausal complexity, nominalization and lexical bundles – show that Luther was the more readable text for its contemporaries. Although Hfa also attempted to produce a readable text, the translators were not as successful, as evidenced by consistently higher lexical density, lower clausal complexity, more frequent use of nominalization, and lower frequency of lexical bundles. These results contribute both to the great amount of research on Luther’s translation, as well as to the discussion of the importance of readability in the choice of Bible translations, as seen in Dewey (2004). Luther is shown here to not only be an enduring work, but one that was more readable and accessible to its contemporaries than a comparable Bible translation from modern times. Critique of Hfa in comparison to Luther can now be further made based on these results, with clear knowledge that Hfa is not as readable for the people of today as Luther was for the people of the 16th century.
By going to the “snouts” of the people he was translating for, looking at how mothers and children speak, Luther’s work is shown to be all the more amazing for its readability, in combination with its longevity as a Bible that is still widely read and used today.


Borin, L. (2002). …and never the twain shall meet?” In L. Borin (Ed.), Parallel Corpora, Parallel Worlds (pp. 2-46). Amsterdam: Rodopi.


Smith, K. (2004). 'I am me, but who are you and what are we?' The translation of personal pronouns and possessive determiners in advertising texts. Multilingua, 23(3), 283-303.


Appendix A

Stuttgart-Tübingen Tagset (Schiller, Teufel and Stöckert, 1995)

The following list shows the different tags used in the calculation of lexical density and the counting of different types of conjunctions, as seen in chapters 4 and 5.

**Content Word Tags:**

NN: normal nouns  
NE: names  
ADJA: attributive adjectives  
ADJD: predicative or adverbial adjectives  
CARD: cardinal numbers  
VAFIN, VAIMP, VVFIN, VVIMP, VMFIN: finite verbs  
VVINF, VVINF, VMINF, VVIZU: infinitives  
VVPP, VMPP, VAPP: perfect participles  
ADV: adverbs

**Conjunction Tags:**  
KOUS: subordinating conjunctions  
KON: coordinating conjunctions
Appendix B

Lexical Density Calculation Program

This program was developed in the Python Programming Language (INSERT REFERENCE HERE). It calculates lexical density based on the Stuttgart-Tübingen Tagset (Schiller, Teufel et. al, 1995). Two versions of this program were developed. The first included all forms of verbs in the calculation of lexical density. The second excluded all forms of modal and auxiliary verbs in the calculation. The output of this program can be seen here:

Figure B.1

![Figure B.1: Lexical Density Stat Program Output](image-url)
The programs themselves are shown below:

**B.1 Lexical Density Calculator with all Verbs**

```python
import sys
import string

#makes nice organized output.
def nicePrint():
    print "---"*20

#allows opening of corpus files for analysis
infile=open(sys.argv[1],"r")
lines = infile.readlines()
infile.close()

#Content word tags and counter, based on Stuttgart Tagset
contentWord=0
totalToken=0
nounCount=0
adjCount=0
cardCount=0
verbCount=0
advCount=0

#counts and calculates lexical density based on tag
for line in lines:
    line=line.strip()
    if line[0] in string.punctuation:
        continue
    else:
        totalToken+=1
taglist=line.split()
    if taglist[1] in content:
        contentWord+=1
    else:
        continue
    if taglist[1]==content[0]:
        nounCount+=1
```
elif taglist[1]==content[1]:
    nounCount+=1
elif taglist[1]==content[2]:
    adjCount+=1
elif taglist[1]==content[3]:
    adjCount+=1
elif taglist[1]==content[4]:
    cardCount+=1
elif taglist[1]==content[5]:
    verbCount+=1
elif taglist[1]==content[6]:
    verbCount+=1
elif taglist[1]==content[7]:
    verbCount+=1
elif taglist[1]==content[8]:
    verbCount+=1
elif taglist[1]==content[9]:
    verbCount+=1
elif taglist[1]==content[10]:
    verbCount+=1
    verbCount+=1
elif taglist[1]==content[12]:
    verbCount+=1
elif taglist[1]==content[13]:
    verbCount+=1
elif taglist[1]==content[14]:
    verbCount+=1
elif taglist[1]==content[15]:
    verbCount+=1
elif taglist[1]==content[16]:
    verbCount+=1
elif taglist[1]==content[17]:
    advCount+=1

#Verbal output for analysis
lexicalDensity=float(contentWord)/totalToken
nicePrint()
print "Your text has", totalToken, "total tokens."
nicePrint()
print "Of those tokens," , contentWord, "are content words."
nicePrint()
print "The lexical density of your text is", lexicalDensity, "."
nicePrint()
B.2 Lexical Density Calculator with Modals and Auxiliaries Excluded

```python
import sys
import string

def nicePrint():
    print "---"*20

infile=open(sys.argv[1],"r")
lines = infile.readlines()
infile.close()

content = ["NN", "NE", "ADJA", "ADJD", "CARD", "VVFIN", "VVIMP", "VVINF", "VVIZU", "VVPP", "ADV"]
contentWord=0
totalToken=0
nounCount=0
adjCount=0
cardCount=0
verbCount=0
advCount=0

for line in lines:
    line=line.strip()
    if line[0] in string.punctuation:
        continue
    else:
        totalToken+=1
taglist=line.split()
        if taglist[1] in content:
            contentWord+=1
nounCount=contentWord/totalToken
adjCount=contentWord
```
contentWord+=1
else:
    continue
if taglist[1]==content[0]:
    nounCount+=1
elif taglist[1]==content[1]:
    nounCount+=1
elif taglist[1]==content[2]:
    adjCount+=1
elif taglist[1]==content[3]:
    adjCount+=1
elif taglist[1]==content[4]:
    cardCount+=1
elif taglist[1]==content[5]:
    verbCount+=1
elif taglist[1]==content[6]:
    verbCount+=1
elif taglist[1]==content[7]:
    verbCount+=1
elif taglist[1]==content[8]:
    verbCount+=1
elif taglist[1]==content[9]:
    verbCount+=1
elif taglist[1]==content[10]:
    advCount+=1

lexicalDensity=float(contentWord)/totalToken
nicePrint()
print "Your text has", totalToken, "total tokens."

nicePrint()
print "Of those tokens," , contentWord, "are content words."
nicePrint()
print "The lexical density of your text is", lexicalDensity, "."
nicePrint()
print "The text contains" , nounCount, "nouns."
nicePrint()
print "The text contains" , adjCount, "adjectives."
nicePrint()
print "The text contains" , cardCount, "cardinal numbers."
nicePrint()
print "The text contains" , verbCount, "verbs."
nicePrint()
print "The text contains" , advCount, "adverbs."
nicePrint()
Appendix C

The *grep* function in UNIX and its use in this project

One of the most important corpus tools used in this project is the *grep* function of UNIX. *Grep* is a tool that searches for patterns within lines of text (Brew and Moens, 2002. p. 36). This powerful tool has several possible functions, as seen below:

- `grep 'text'` find all lines containing the word \text
- `grep '^text'` find all lines beginning with the word \text
- `grep 'text$'` find all lines ending in the word \text
- `grep '[0-9]'` find lines containing any number
- `grep '[A-Z]'` find lines containing any uppercase letter
- `grep '^[A-Z]'` find lines starting with an uppercase
- `grep '[a-z]$'` find lines ending with a lowercase
- `grep '[aeiouAEIOU]'` find lines with a vowel
- `grep '[^aeiouAEIOU]$'` find lines ending with a consonant (i.e. not a vowel)
- `grep -i '[aeiou]'` find lines ending with a vowel (ignore case)
- `grep -v 'text'` print all lines except those that contain \text
- `grep -v 'text$'` print all lines except the ones that end in \text

(p. 35-36)

Lines containing each suffix were found with *grep* and sorted by frequency. A sample search using *grep* in Unix would look like this:

```
cbs4[4]: grep 'nis$' hfant_freq
    36 gefaengnis
    14 finsternis
    12 gleichnis
    6 begraubnis
    3 geheimnis
    3 erkenntnis
    2 verstaendnis
    2 kenntnis
    1 zugestaendnis
    1 zeugnis
    1 verhaeltinis
    1 erlaubnis
    1 ereignis
```
VITA

Nathan L. Shrefler

EDUCATION:

Ph.D. German Applied Linguistics, August 2008 GPA: 3.78
The Pennsylvania State University, University Park, PA
Advisor: B. Richard Page, PhD, Associate Professor of German and Linguistics
Dissertation: Readability and German Bibles

B.A. German with Honors and Distinction in German, December 2001 GPA: 3.76
The Pennsylvania State University, University Park, PA
Advisor: B. Richard Page, PhD, Associate Professor of German and Linguistics
Honors Thesis: The Legacy of the Fall of the Berlin Wall and the Reunification of Germany.

Penn State Study Abroad Program, September 2000 – July 2001
Philipps Universität, Marburg an der Lahn, Germany

TEACHING EXPERIENCE:

Teaching Assistant, 2004 – 2008 The Pennsylvania State University, University Park, PA
Supervisor, Hülya Yilmaz, PhD, Senior Lecturer in German

Graduate Fellow, 2003 – 2004 The Pennsylvania State University, University Park, PA
Supervisor: Dr. Ulrich Andersch

Fulbright Teaching Assistant, 2002 – 2003 Freiherr vom Stein Schule, Gladenbach, Germany
Supervisor: Dennis Guth, Director of Personnel

HONORS:

Graduate Research Exhibition, 1st Place Poster, Arts and Humanities, 2008

Dissertation Fellow, The Institute for Arts and Humanities, The Pennsylvania State University, 2007

Thompson Scholarship, The Pennsylvania State University, 2004 – 2007

Graduate Fellowship, The Pennsylvania State University, 2003

Fulbright Travel Grant, The German/American Fulbright Commission, 2002

Fulbright Teaching Assistantship, Pädagogischer Austauschdienst, Germany, 2002-2003

Phi Beta Kappa Society, 2001


Schreyer Honors College Travel Grant, The Pennsylvania State University, 2000

National Society of Collegiate Scholars, 1999

Phi Eta Sigma National Honor Society, 1998

Schreyer Honors College Academic Excellence Scholarship, 1997 – 2001

PROFESSIONAL AFFILIATIONS

The Institute for Arts and Humanities, The Pennsylvania State University, 2007
The Modern Language Association, 2007-2008