THE SYNCHRONY AND DIACHRONY OF BOSNIAN-CROATIAN-SERBIAN
ADJECTIVAL LONG-FORM ALLOMORPHY (ALFA)

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy
in the Graduate School of The Ohio State University

By
James Joshua Pennington

Graduate Program in Slavic and East European Languages and Literatures

The Ohio State University
2011

Dissertation Committee:
Brian D. Joseph, Advisor
Daniel E. Collins
Charles E. Gribble
ABSTRACT

In Bosnian-Croatian-Serbian (BCS), the genitive (G) and dative/locative (DL) cases exhibit adjectival long-form allomorphy (ALFA). The genitive –og ~ -oga and the DL – om ~ -ome ~ -omu stand in free variation, inasmuch as when one form is substituted for another the truth value of an utterance remains unchanged. Some sociolinguists (particularly Milroy & Gordon 2003) have held the view that while morphosyntactic free variation is possible it is much harder to analyze from the sociolinguistic perspective. While it is undoubtedly true that phonological variables occur much more frequently than so-called “higher-level” variables, many valuable contributions to sociolinguistics (Labov 1969, Wolfram 1969, Rickford et al. 1991, Mufwene 1998, Poplack and Tagliamonte 2001, and many others) have focused on morphosyntactic variables, and consequently have revealed important patterns of morphosyntactic variation in speech communities. Surprisingly, this sort of research has rarely been conducted on the South Slavic languages, especially BCS, which exhibits an incredible amount of morphosyntactic variation, given that Bosnian, Croatian, and Serbian have shared dialect continua for centuries and are mutually intelligible, but now stand at odds with one another due to
changes in language policy stemming from the breakup of the Former Yugoslavia (Greenberg 2004).

This dissertation is a fundamental rejection of the notion that is still being proffered (at least in Croatian grammars, cf. Barić 2007, Raguž 2010) that there are contextually based rules or tendencies for ALFA usage in Contemporary BCS. While previous grammars may have stated the facts accurately for their synchronic period of BCS, many contemporary grammars have essentially “piggy-backed” on these prior grammars (e.g Shirokov & Gudkov 1977, Browne & Alt 2004), which viewed these ALFAs as adding a sense of “definiteness”, and subsequently have neglected the possibility of language change.

Alexander (2006) is the first to propose that ALFAs are more common in Croatian than in Bosnian or Serbian. My sociolinguistic questionnaire and analysis of the Croatian Language Repository, the Croatian National Corpus, and Wikipedia have all verified that ALFAs are more common in Croatian, although they can be heard rarely in Serbian (only for possessives and demonstratives), and slightly more frequently in Bosnian.

Furthermore, the empirical studies conducted in this investigation have informed the diachronic origin of ALFAs. It can certainly be said that ALFAs occur infrequently with descriptive adjectives, but exhibit strong patterns of variation for possessives and demonstratives. Moreover, for monosyllabic possessives and demonstratives (which are still monosyllabic after the addition of G and DL base forms –og and –om), when a preposition precedes the adjectival phrase the ALFAs are overwhelmingly preferred. I
link this tendency to pronominal prepositional phrases which require long forms, e.g. *o meni, but *o mi. This can be viewed as a “two-syllable” (or “two-footed”) constraint, which then spread to all prepositional phrases in the language via analogy. This “two-syllable” rule can be viewed as a restatement of “Wackernagel’s position”, which can be either prosodic or syntactic (Serbian vs. Croatian, for example), in syllabic terms, although it is possible that this process could have begun on the moric level *o vās, but *o vas. In the absence of a noun this element either had already took on, or developed an anaphoric relationship to a following noun, and thus was reinterpreted as a marker of definiteness. Subsequent analogies spread ALFAs to the descriptive, although the “two-syllable” constraint would never have been violated for descriptive, given that all descriptive roots after the addition of ALFA bases would have at minimum two syllables already.

Finally, the lack of ALFAs in Western Štokavian (one of the historical predecessors of modern-day Croatian) pre-19th century and their early attestation in 15th century Eastern Štokavian (the historical predecessor of modern-day Serbian) points to a Serbian import of ALFAs to Croatian lands. Ivić (1972) has demonstrated that the political and economic repercussions of the Ottoman occupation of Serbia led many Serbs to seek asylum with their Christian neighbors. As repeated migrations of Serbs brought to Croatian dialects characteristic Serbian features, such as: the neo-Štokavian stress retraction, the da complementizer, etc., the linguistic conformance behind the Vienna Literary Agreement of 1850 would have viewed the presence of these features as “unifying” elements.
Therefore, the statement made by Greenberg (2004) that the ALFAs were reintroduced as “puristic” elements in Croatian post-1991 can be considered an ironic twist in Croatian language policy, inasmuch as, from the diachronic perspective, ALFAs have been shown to be characteristic of 15th century Serbian, and then later Croatian.
DEDICATION

I dedicate the following work to my father and mother. My father, who, despite having spent the last year in poor health, has continually inspired me with his generally upbeat nature, constantly exhorting me to the “finish line”, so to speak.

My mother’s reassuring voice has been a calming force during times of trepidation. Like the ideal mother, she effortlessly juggles her own difficulties with her children’s, who acknowledge this by striving to be successful in remittance for a wonderful childhood. My gratitude is boundless.
I would like to finally attempt to express in words the sincere appreciation I have come to have for Professor Brian D. Joseph’s mentorship, personally and professionally. I have never known anyone quite like Brian (as he asks his students to call him), who holds such an intimidating record of scholarly and professional accomplishments, but whose humility and down-to-Earth nature endear him to his peers and students. Like a great craftsman (a definite necessity in my case), Brian has whittled my scholarly philosophy from a disconnected block of linguistic theory to something of a more refined shape. Now, like an Orthodox icon waiting to be brought to vivid life by Rublev, I believe my scholarly career will be brilliantly painted by following in the footsteps of Brian D. Joseph.

Before Brian became my advisor, however, there was Dr. Charles E. Gribble, or “Gribbs”, as many of us graduate students used to affectionately call him (never to his face, of course!). Possessing truly encyclopedic knowledge, Dr. Gribble is the one of the few remaining Slavic linguists of the Jakobsonian (he studied under Jakobson) philological type. I try my best to emulate Dr. Gribble’s passion for knowledge and attention to the smallest details, although I know this is probably well out of my reach.
Dr. Daniel E. Collins has always been a great sounding board for my sometimes misguided ideas. While he has constantly supported my at times overly inquisitive bent, he has viewed some of my less successful works as being brave attempts. While many junior scholars might take offense to such an evaluation, I understand Dr. Collins’ sentiment. He sees potential, but recognizes flaw, and for this I am truly grateful. It was never enough for me to present on safe topics, but to explore and challenge core assumptions about Slavic diachronic linguistics. Although a few of these works can ultimately be judged failures, I thank Dr. Collins for seeing potential in them. It is up to me solely to eradicate their flaws.

Finally, I would like to thank the wonderful people of Bosnia, Croatia, and Serbia for taking their time and effort to fill out my sociolinguistic questionnaires. It was through these questionnaires that I have made some true friends—genuinely good people who were delighted to see they are not forgotten in the West.
VITA

September 26, 1981...............................Born—Portsmouth, Ohio

June, 2003.................................B.A. Russian, Ohio University

March, 2006.................................M.A. Slavic Languages and Literatures

2004-Present...........................Graduate Teaching Associate,

The Department of Slavic and Eastern

European Languages and Literatures

The Ohio State University

PUBLICATIONS


FIELDS OF STUDY

Major Field: Slavic and East European Languages and Literatures

Slavic Linguistics

x
TABLE OF CONTENTS

Abstract........................................................................................................................................ iii
Dedication....................................................................................................................................... vii
Acknowledgments........................................................................................................................... viii
Vita................................................................................................................................................ x
List of Tables................................................................................................................................... xiv
List of Figures.................................................................................................................................. xv

Chapters:
1 Introduction................................................................................................................................. 1
2 Morphosyntax in Light of Sociolinguistics............................................................................... 7
   2.1 Views of Morphosyntax as "Less" Socially Indexed......................................................... 8
      2.1.1 Positing the Linguistic Variable.............................................................................. 9
   2.2 Statistical Representativeness and "Leaky" Systems..................................................... 14
3 Historiography of BCS Adjectival Long-form Allomorphy (ALFA)................................. 25
   3.1 Serbian Grammars............................................................................................................. 28
      3.1.1 Karadžić, Vuk Stefanović. (1824).......................................................................28
      3.1.2 Hraste, Mate and Sreten Živković. (1949)......................................................... 29
      3.1.3 Belić, Aleksandar. (1962).................................................................................. 32
      3.1.4 Stevanović, Mihailo (1964)............................................................................... 32
   3.2 Non-native Grammars of Serbo-Croatian/BCS............................................................... 34
      3.2.1 Shirokova and Gudkov, eds. (1977).................................................................... 34
      3.2.2 Browne, Wayles and Theresa Alt. (2004)........................................................... 35
      3.2.3 Browne, Wayles. (2006).................................................................................... 38
      3.2.4 Alexander, Ronelle. (2006)................................................................................ 38
3.3 Croatian Grammars………………………………………………………….41
  3.3.1 Kašić, Bartol. (1604).................................................................41
  3.3.2 Ardelio della Bella. (1728).........................................................42
  3.3.3 Pacel, Vinko. (1860).................................................................42
  3.3.4 Maretić, Tomislav (1899).......................................................43
  3.3.5 Barić, Eugenija (2005) and Dragutin Raguž (2010).............45

4 An ALFA Sociolinguistic Questionnaire..............................................49
  4.1 Quantitative and Qualitative Aspects...........................................50
    4.1.1 Quantitative Aspects of Sociolinguistic Fieldwork..................51
    4.1.2 Qualitative Aspects..........................................................52
    4.1.3 The Golden Horseshoe Project............................................53
  4.2 The ALFA Questionnaire..........................................................54
    4.2.1 The Structure of the Questionnaire......................................55
    4.2.2 The Linguistic Criteria of the Questionnaire..........................57
    4.2.3 Native Speaker Assessment of Sentences...............................58
  4.3 Removing the Observer's Paradox from My Sociolinguistic Survey....59
    4.3.1 The Use of Misdirection....................................................59
  4.4 Results of the ALFA Questionnaire............................................60
    4.4.1 Breakdown of ALFA Usage for non-Croatian Respondents.........61
    4.4.2 Results for Croats living in Croatia.....................................63
    4.4.3 ALFA Usage for Croatian Men vs. Women.............................64
  4.5 Concluding Remarks.....................................................................66

5 An Empirical Study of ALFA in Bosnian and Croatian..........................68
  5.1 Materials Used in This Study....................................................69
  5.2 Methodology/Query Criteria.....................................................70
  5.3 Results of Corpus-based Analysis...............................................71
    5.3.1 Total tokens for –om, -ome, -omu in the Corpus of Bosnian Texts.71
    5.3.2 The distribution of G –og/–oga in Bosnian.............................73
    5.3.3 ALFA Distribution in the Croatian Online Corpora....................74
    5.3.4 Distribution of G –og/–oga in the OCBT, the CLR, and the CRC...78
    5.3.5 Distribution of DL ALFAs in the Bosnian and Croatian Wikipedia
        Corpora..................................................................................79
    5.3.6 Distribution of Croatian ALFAs according to Adjectival Class.....82
    5.3.7 Consecutive ALFAs in Croatian.............................................89
    5.3.8 ALFAs and Preceding Prepositions........................................95
    5.3.9 The Somewhat Frozen Phrase zbog toga „because of, because”….100
    5.3.10 The Distribution of DL ALFAs –om, -ome, -omu when Preceded
          by a Preposition................................................................101
    5.3.11 Distribution of ALFAs according to Type of Literature..........105
  5.4 Conclusions of Corpus Analysis of G and DL ALFA........................108

6 The Diachronic Aspect of BCS ALFA..................................................111
6.1 Proposed Explanations for the Rise of BCS ALFA ..............................111

6.2 The Rise of Non-Etymological ALFA –e via Analogical Leveling to
the Nominal Paradigm ........................................................................116

6.3 An Integrated Hypothesis for the Rise of Unexpected –e and its Spread to
Croatian Dialects ................................................................................120
   6.3.1 New Hypothesis for the Rise of Unexpected ALFA –e in Eastern
Štokavian ..............................................................................................123
   6.3.2 The Spread of ALFA –e to Croatian Dialects ...............................131

7 Conclusions .......................................................................................135

Appendices:
A A Chronology of Croatian/Illyrian Grammars .................................141
B The Official/Academic Usage of ALFAs ...........................................147
C Frequency List of Individual Adjectives in CNC with DL ALFAs .........150
D Questionnaires ..................................................................................153
E Data Tables for Figures from Chapter 5 .............................................164
F Dialect Map of BCS Speech ...............................................................173
List of References ...............................................................................175
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1. ALFA for descriptive adjectives in Karadžić (1824)</td>
<td>29</td>
</tr>
<tr>
<td>3.2. ALFA for possessives and demonstrative adjectives in Karadžić (1824)</td>
<td>29</td>
</tr>
<tr>
<td>3.3. ALFA as listed in Belić (1962)</td>
<td>32</td>
</tr>
<tr>
<td>4.1 Respondent breakdown according to ethnicity and gender</td>
<td>60</td>
</tr>
<tr>
<td>4.2 ALFA results for non-Croatian respondents</td>
<td>61</td>
</tr>
<tr>
<td>4.3 Results for Croats living in Croatia</td>
<td>64</td>
</tr>
<tr>
<td>4.4 Results for male vs. female ALFA usage</td>
<td>66</td>
</tr>
<tr>
<td>5.1 DL ALFAs in the Oslo Corpus of Bosnian Texts</td>
<td>72</td>
</tr>
<tr>
<td>5.2 Frequency distribution of most common adjectives with DL ALFA</td>
<td>73</td>
</tr>
<tr>
<td>5.3 G ALFA in the Oslo Corpus of Bosnian Texts</td>
<td>74</td>
</tr>
<tr>
<td>5.4 ALFA distribution in the CLR</td>
<td>75</td>
</tr>
<tr>
<td>5.6 Breakdown of text type in the CNC</td>
<td>76</td>
</tr>
<tr>
<td>5.7 Distribution of ALFAs –og/–oga in the OCBT</td>
<td>78</td>
</tr>
<tr>
<td>5.8 Distribution of ALFAs –og/–oga in the CLR</td>
<td>79</td>
</tr>
<tr>
<td>5.9 Distribution of ALFAs –og/–oga in the CNC</td>
<td>79</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.7</td>
<td>Browne &amp; Alt’s (2004) long and short adjectival paradigms</td>
<td>36</td>
</tr>
<tr>
<td>5.1</td>
<td>71,449 ALFs: –om: 66159 (93%) ; –ome: 4810 (7%) ; –omu: 480 (.07%)</td>
<td>80</td>
</tr>
<tr>
<td>5.2</td>
<td>125,984 ALFs: -om: 115,811 (92%) ; -ome: 8,934 (7%) ; –omu: 1,239 (1%)</td>
<td>81</td>
</tr>
<tr>
<td>5.3</td>
<td>Distribution of Croatian DL ALFAs for possessive adjectives</td>
<td>83</td>
</tr>
<tr>
<td>5.4</td>
<td>Distribution of Croatian G ALFAs for possessive adjectives</td>
<td>84</td>
</tr>
<tr>
<td>5.5</td>
<td>Distribution of Croatian DL ALFAs for demonstrative adjectives</td>
<td>85</td>
</tr>
<tr>
<td>5.6</td>
<td>Distribution of Croatian G ALFAs for demonstrative adjectives</td>
<td>86</td>
</tr>
<tr>
<td>5.7</td>
<td>Distribution of Croatian G and DL for other non-descriptives</td>
<td>87</td>
</tr>
<tr>
<td>5.8</td>
<td>Distribution of Croatian G and DL ALFAs for descriptive adjectives</td>
<td>88</td>
</tr>
<tr>
<td>5.9</td>
<td>ALFAs for two consecutive DL adjectives in Croatian</td>
<td>90</td>
</tr>
<tr>
<td>5.10</td>
<td>ALFAs for three consecutive DL adjectives in Croatian</td>
<td>92</td>
</tr>
<tr>
<td>5.11</td>
<td>ALFAs for two consecutive G adjectives in Croatian</td>
<td>93</td>
</tr>
<tr>
<td>5.12</td>
<td>ALFAs for three consecutive G adjectives in Croatia</td>
<td>94</td>
</tr>
<tr>
<td>5.13</td>
<td>G and DL ALFAs when preceded by a preposition</td>
<td>96</td>
</tr>
<tr>
<td>5.14</td>
<td>G ALFAs for possessive adjectives with a preceding preposition</td>
<td>97</td>
</tr>
<tr>
<td>5.15</td>
<td>G ALFAs for demonstrative adjectives with a preceding preposition</td>
<td>98</td>
</tr>
<tr>
<td>5.16</td>
<td>G ALFAs for other non-descriptive adjectives with a preceding preposition</td>
<td>99</td>
</tr>
<tr>
<td>5.17</td>
<td>G ALFAs for descriptive adjectives with a preceding preposition</td>
<td>100</td>
</tr>
<tr>
<td>5.18</td>
<td>DL ALFAs for other possessive adjectives with a preceding preposition</td>
<td>102</td>
</tr>
<tr>
<td>5.19</td>
<td>DL ALFAs for demonstrative adjectives with a preceding preposition</td>
<td>103</td>
</tr>
</tbody>
</table>
5.20 DL ALFAs for other non-descriptive adjectives with a prec. preposition.......104
5.21 DL ALFAs for descriptive adjectives with a preceding preposition.............105
5.22 Distribution of ALFAs in croatian subtitled foreign films and shows.............107
CHAPTER 1

INTRODUCTION

In the South Slavic dialects referred to collectively as Bosnian-Croatian-Serbian (hereafter BCS), speakers can choose from the following set of three possible inflectional morphemes (adjectival long-form allomorphy—hereafter ALFA) to mark adjectives in the masculine and neuter forms for the dative and locative (hereafter DL) cases: -om, -ome, -omu. Similarly, in the genitive case (hereafter G), speakers can choose from either –og and –oga. For example:

(1a) Govorim o dobrom-DL sinu.
'I speak about the good son’.

(1b) Govorim o dobrome-DL sinu.
'I speak about the good son’.

(1c) Govorim o dobromu-DL sinu.
'I speak about the good son’.

1 Following now standard Western practice, throughout this work I refer to the language in question as BCS (Bosnian-Croatian-Serbian), unless there is cause to single out a specific variety (cf. Chapter 5 on my corpus-based Investigation, for instance). BCS represents the successor language to the official language of Former Yugoslavia, which had at least three different titles during the Communist period, e.g.: Srpkohrvatski/hrvatskosrpski, hrvatski ili srpski (this is the title of the 1988 Brozović and Ivić grammar).

2 Where discussed as a general topic, read: “allomorphy”; where discussed in the context of the genitive or dative/locative I use the plural form ALFAs, in the sense “adjectival long-form allomorphs”.

3 There also exist so-called short forms for the G, but these forms do not have ALFA variants (e.g. dobra psa ‘of the good dog’).
(2a) Vidim malog-G psa.
‘I see a small dog’.

(2b) Vidim maloga-G psa.\(^4\)
‘I see a small dog.’

All five examples provided above are illustrative of what can be heard in various parts of the former Yugoslavan republics of Bosnia, Croatia, Montenegro, and Serbia.\(^5\) At times, a speaker’s choice of one ALFA over another may appear haphazard, as Ronelle Alexander has stated: “[the ALFAs] can be substituted at the speaker’s choice…” and “…there is no consistent difference in meaning among them, other than the rather fluid one of style” (Alexander 2006: 53).

In each of the above examples for the sentence “I speak about the good son”, there are exactly three propositions subject to various truth conditions:

1) There is a son.

2) This son is good.

3) I speak about him.

The addition of the long-form ALFAs does not render any of the propositions false, and therefore, from a strict semantic perspective, they can be viewed as semantic equivalents (although pragmatic and social motivations could certainly be present, as this is discussed in greater detail later).

While the ending -om is by far the most common of the three endings (which are enumerated above alphabetically relative to frequency of occurrence), the other

\(^4\) BCS also has the “longer” form svojega, but this is irrelevant to my study. The critical issue addressed here is not the longer internal syllable, which is the historical invariable form, but the addition of –a, which, like the –e of the DL, has no clear etymology. The origins of –oga are discussed in Chapter 7.

\(^5\) For real examples derived from grammars and corpora, see Chapters 4, 5, and 6.
two occur much less frequently and pose analytic problems as to their use and 
distribution. In that way, they present a challenge to linguists studying the BCS 
dialects somewhat like the variation in complementation involving infinitives and 
finite clauses. For example, in BCS there are two ways to express infinitival 
constructions, as in English *to read, to speak*, etc.: 1) using an infinitival verb form, 
e.g.: *piti* ‘to drink’ or 2) using the subordinating particle *da* + a present tense verb 
form, e.g.: *da pijem* ‘to drink’.

A parallel example on the lexical level is the 
distribution of *hleb* ‘bread’ and *kruh* 'bread'. While these variants are semantic 
equivalents, they do not share an equal distribution across the BCS dialectal territory. 
*Da* + present for complement clauses and *hleb* belong to the dialect of BCS known as 
'ekavski' (mostly Serbian, although *da* + present can be heard also in Croatia and 
Bosnia), while infinitival forms and *kruh* belong to 'ijekavski' (mostly Bosnian and 
Croatian although they can be heard in Serbia and Montenegro). In other words, these 
features are typically viewed as being regional variants, which are used by different 
social groups, although they can be and often times are shared along dialect continua.

I argue here that the DL ALFAs –*om*, –*e*, and –*u*, are another such case of 
regional variants, and can be viewed as being distributed according to ethnicity. In 
fact, my central claim is that the type of ALFA distribution present in a dialect of BCS 
is directly tied to national identity (the intersection of ethnicity and geography). 
However, let us delay this discussion until I have presented the full scope of BCS 
ALFA.

A related issue, which has yet to be resolved, is the origin of the unexpected –*e* 
of DL –*ome* (cf. Common Slavic D *–omu*, L *–omi*; also G *-ogo*). Attempts by 

---

Marković (1964) to explain the rise of –e as a result of phonetic processes or by Belić (1962) to treat -e as a sort of particle (unexplained in his account) appear ad hoc when they are juxtaposed with a view that considers a variety of factors, including analogical leveling, thus taking in not only morphological considerations, but also syntactic and prosodic. I argue later that these diachronic facts inform potential motivations for the current dialect distribution of the ALFAs.

On the whole, this study attempts to solve two questions: (1) In what ways are BCS ALFAs distributed synchronically?; and 2) What is the origin of BCS ALFA? In order to answer these questions in the fullest scope possible, I have taken a synchrony/diachrony approach to ALFA variation, in the spirit of Joseph (1983),7 to trace the development of ALFA from its origins to its modern situation. Below, the reader is presented with a chapter-by-chapter preview of how my synchrony/diachrony approach unfolds.

In Chapter 2, my investigation begins with an overview of some of the major works on morphosyntactic variation, including Wolfram (1969), Rickford et al. (1991), Mufwene (1998), Rickford (1999), Poplack and Tagliamonte (2001), and Walker (2000), which all share a central commonality: the impact of social identity on variation. These works justify looking at BCS ALFA from a sociolinguistic perspective. However, some sociolinguists (e.g. Milroy & Milroy 2002, et al.) have been skeptical whether morphosyntactic variation can be genuinely sociolinguistic in nature. Essentially, these claims are based on the idea that higher frequency of form equals higher social indexing. However, I present counter-arguments that suggest

---

7 Joseph found that the synchronic lack of infinitival constructions in the Balkan languages is related to many diachronic factors, including language contact with Greek.
morphosyntactic variation can, in fact, be sociological in nature, and should be studied sociolinguistically, where feasible.

In Chapter 3, I present a historiography of the literature that has been written on the DL adjectival paradigms in Bosnian, Croatian, and Serbian. For Serbian, I begin with Vuk Karadžić’s *Kleine Serbische Grammatik* (1824) and end with Ronelle Alexander’s *Bosnian-Croatian-Serbian: A Grammar with Sociolinguistic Commentary* (2006). For Croatian, I go back to the first *Illyrian* grammar ever printed, Bartolo Kašić’s *Osnove ilirskoga jezika* (1604), and end with Dragutin Raguž’s *Praktična hrvatska gramatika* (2010). For Bosnian, which only gained status as an official language, separate from Croatian and Serbian, in 1994, I consult Midhat Ridañović’s *Bosnian Grammar* (2010).

Chapter 4 is where the “synchronic” aspect takes over, so to speak. I present the results of a sociolinguistic questionnaire I distributed in the summer of 2009, while traveling through Bosnia, Croatia, and Serbia. The questionnaire inevitably verifies claims that ALFAs are more Croatian, thus excluding Serbian from the remainder of this study. Moreover, as no unified linguistic explanation exists for the distribution of ALFA in BCS, I look at social variables, particularly sex and nationality, which have proven to be categories that have implications for the usage of ALFAs in Croatia.

---

8 Up until the second half of the 19th Century, this term was used widely for the area called modern-day “Croatia”, and it was popularized by Croats in the 19th Century during a period of language revival. See Appendix A for a list of Illyrian/Croatian grammars in chronological order. The term “Illyrian” seems to have died out, as might be expected, with the signing of the Vienna Literary Agreement in 1850, unifying the dialects of Serbia and Croatia. Between 1850-1860, “Illyrian” was slowly replaced by “Croatian” in the naming of the grammars.
In Chapter 5, I present my corpus work of ALFA in Bosnian and Croatian. By feeding the Bosnian and Croatian Wikipedia into the *R Project for Statistical Computing*, and by utilizing online national corpora for Croatian and Bosnian, I analyze the distribution of ALFA in the various environments proposed in previous accounts.

Chapter 6 is an examination of prior scholarship on the origins of ALFA variation in BCS. I challenge this scholarship by examining ALFA variation drawing on evidence from historical dialectology of the modern-day BCS territories. Furthermore, employing analogical theory and continuing the widely held belief that ALFAs were originally markers of definiteness, I propose a theory that views ALFAs as “nominal clitics” in the absence of a following noun.

Finally, in Chapter 7, I present a synthesis of the problems researched in this dissertation. Incorporating the synchronic and diachronic aspects of my research, I offer new conclusions drawn from a careful examination of available data and previous accounts of ALFA usage in BCS grammars.

---

9 The Serbian Wikipedia data was unnecessary, as ALFAs are strictly avoided in forms other than possessive and demonstrative adjectives.
CHAPTER 2

MORPHOSYNTAX IN LIGHT OF SOCIOLINGUISTICS

Since the publication of William Labov’s seminal work “The social motivation of sound change” (Labov 1963), linguists have recognized and developed further the notion that sound change is often tied to social factors. However, this has resulted in sociolinguistics becoming dominated by empirical studies of so-called "lower-level"\(^{10}\) variation, e.g. phonology, vowel shifts, etc. to the exclusion of “higher-level” variation, e.g. morphology and syntax. In fact, in Labov’s *The Principles of Linguistic Change: Social Factors* (2001), there are exactly zero pages dealing with morphological or syntactic variation. In his *Principles of Linguistic Change: Internal Factors* (1994), Labov does admit that a complete view of linguistic change “would include the study of semantic change, syntactic development, changes in word order, and the grammaticalization of morphological elements as well as phonological and phonetic change” (Labov 1994: 30). Although *Internal Factors* and focus strictly on the intersection of sound change and social factors, Labov was, in fact, the first to extend variationist studies to morphology in his work on copula deletion in English (Labov

\(^{10}\)The terms “low-level” and “high-level” as descriptions of linguistic phenomena stems from the general (although not necessarily useful) notion that elements that have less meaning are considered “low-level”, e.g. phonological features, as opposed to “higher-level” phenomena, e.g. morphology and syntax.
Through a meticulous empirical analysis, Labov was able to conclude that there was in fact great variation in the environments for copula-deletion in African-American Vernacular English (AAVE), and that each type of variation was indicative of a subgroup of AAVE speakers. Moreover, his findings showed that ethnicity could be an underlying factor in morphosyntactic variation (inasmuch as AAVE patterns varied from standard American patterns). Despite this groundbreaking work by Labov and subsequent works on AAVE, including: Wolfram (1969), Rickford et al. (1991), Mufwene (1998), Poplack and Tagliamonte (2001), there have been a fair number of criticisms proffered by sociolinguists that higher-level variation is not genuinely sociolinguistic in nature, and these criticisms may have deterred sociolinguists from focusing on higher-level phenomena from the variationist perspective. Section 2.1 gives an overview of these major criticisms.

2.1 Views of Morphosyntax as “Less” Socially Indexed

Many sociolinguists, including Hudson (1980: 157), Romaine (1984), and Garcia (1985) have suggested a sociolinguistic split between so-called “higher-level” variation (syntax and morphology, etc.) and “lower-level” variation (phonetics, prosody, etc.) with "higher-level" variation being considered intrinsically non-sociolinguistic in nature. Instead, they believe morphosyntax should be relegated to the realm of pragmatics, inasmuch as “[higher-level variation—JP] seldom, if ever, serves the function of distinguishing social groups in the way that "classic" phonological and morphological variables do” (Winford 1996:188). Furthermore, Garcia (1985:198) has offered the following as an even tougher critique of
sociolinguistically motivated higher-level variation: "Are there, can there in fact be 'different ways of saying the same thing' among those units whose function is, precisely, to enable speakers to say different things?" Winford (1996: 188) sees the answer to this question as an obvious "no". However, one central question that sociolinguists have continued to disagree on is the basis for positing a linguistic variable.

2.1.1 Positing the Linguistic Variable¹¹

The linguistic variable has been classically defined as a “structural unit with two or more variants that co-vary with internal or external variables” (Winford 1996:184). In his article “The Problem of Syntactic Variation” (1996), Winford discusses the redefining of the concept of the linguistic variable beginning with William Labov, who, in a straightforward manner, defines the linguistic variable as “alternative ways of saying the same thing” and as having “the same truth value” (Labov 1978:2). Weiner and Labov (1983:22) update this definition slightly by stating that linguistic variables “should be truth-conditionally equivalent and used on the whole to refer to the same set of affairs”. However, this definition is little more than a paraphrase of Labov’s earlier statement. This has led some linguists (Romaine 1984, Lyons 1977) to take issue with the idea that truth-value constancy can guarantee descriptive

¹¹ The latest work by Tagliamonte (2006), is considered part of the ‘canon’ of sociolinguistic literature; however this work only touches on the ongoing debate between two different views on the linguistic variable, with one camp viewing the linguistic variable as more or less a phonological phenomenon, and the other seeing the linguistic variable as something that occurs in all levels. I chose Milroy and Gordon because their 2002 book is still highly influential, and they synthesize the harshest criticisms of the social basis for morphosyntactic variation.
constancy. Lyons gives the following examples to counter Weiner and Labov (Lyons 1977: 202):

a) *John is a fool.*

b) *John is a linguist.*

While these two statements can be substituted for one another without changing their truth values (a) that John is a fool and b) that he is also a linguist), they obviously do not have equal descriptive value. Essentially, the argument is that fools do not equal linguists. However, if instead variable a) was *John specializes in the structure of language,* or variable b) *John is dim-witted,* then both truth values and descriptive values would be equal. Lyons’ examples a) and b) have led sociolinguists to more or less settle on the following revised definition of the linguistic variable offered by Lyons (1977: 202):

> Our criterion for sameness and for difference of sense will be made…directly dependent upon the descriptive meaning of utterances. Two or more expressions will be defined to have the same sense (i.e. to be synonymous) over a certain range of utterances if and only if they are substitutable in the utterances without affecting their descriptive meaning…If the utterances are such that they have a determinate truth value, constancy of descriptive meaning will guarantee constancy of truth value. The converse, however, does not hold, for the substitution of one expression for another may change the descriptive meaning of an utterance without thereby altering the truth value.

Essentially, Weiner and Labov (1983:32) were on the right track with truth conditions and descriptive value, only they failed to distinguish which of the two criteria was essential to semantic equivalence or synonymy.

Despite clearing up problems with earlier definitions of semantic equivalence, Lyons leaves out an essential piece of the puzzle of linguistic variation: pragmatics. According to Winford (1996), variation that can be explained by pragmatic
constraints, of course, should not be considered to be a set of semantically equivalent variables, inasmuch as variables that are non-referentially semantically equivalent are given additional semantic coloring according to the pragmatic context. Moreover, such a definition of linguistic variation benefits from the inclusion of the concepts of I-meaning (Intention) and S-meaning (Significance), and a third type of meaning, L-meaning (Literal) introduced by Geis (1994). According to Winford (1996:189), while intention and significance are two pragmatic strategies which rule out semantic equivalence, literal meaning, conversely, is the concept that must remain constant among linguistic variables in order for them to be considered for variationist inquiry. In this view I-meaning and S-meaning can be viewed as a two-tiered filter, leading down to L-meaning. If a purported variable is filtered down to L-meaning, then it can be considered a true “linguistic variable”.

Returning to the topic of this dissertation, BCS ALFA, it appears the long forms –ome and -omu are filtered, so to speak, through this sieve of pragmatic possibilities, ruling out I-meaning and S-meaning, inasmuch as the addition of the ALFAs does not change the truth or descriptive values of a given utterance (cf. exx. from Chapter 1). Moreover, grammars of the last half-century offer conflicting views of ALFA usage (suggesting uncertainty over their exact function), and native speakers have mixed judgments about their distribution (cf. Chapter 4). Furthermore, though once thought to mark definiteness on adjectives, long forms no longer serve this function outside the nominative (Alexander 2006: 53). With more evidence “for”, and little or no evidence “against”, semantic equivalency, BCS ALFA appears to be a set of variables that share the same “L-meaning”. However, there may be pragmatic
issues at play (to be discussed later), which could reopen the possibility that ALFA usage is an expression of I-meaning.

The view that ALFA choice is an expression of I-meaning would imply a pragmatic strategy that differentiates the variable, “higher-level”, morphemes in a “stylistic” sense, to an effect which would be salient to members of a speech community in an extra-linguistic (sociolinguistic) sense. This idea has been stated before by Lavandera (1978), who views sociolinguistic distribution of variation as implying competition among variants in various social groups, which ultimately serves to distinguish these social groups from one another. However, Lavandera suggests that this is not the case with some studies that have claimed to present higher level social differentiation. For example, Lavandera criticizes Labov and Weiner’s work on active/passive constructions in English (1977), which was unsuccessful in finding a sociological reason for variable usage among speakers, and instead the authors settled on defining the phenomenon simply as a “well-established variable in English” (inasmuch as each variable entails the same truth value) (Labov & Weiner 1977: 11). Puzzled by this seemingly anti-climactic conclusion, Lavandera goes on to ask “…are we in the presence of a new concept, the linguistic variable which carries no meaning, which therefore does not signify by means of relative frequencies, and which is different from the sociolinguistic variable? The concept of ‘linguistic’ and ‘sociolinguistic’ is the same one” (Lavandera 1978: 178).

Despite Lavendera’s criticisms of Labov and Weiner (1977) and of higher level variation, generally, I demonstrate in this dissertation that, while there have been many attempts by BCS grammarians and linguists to pin down the distribution of
ALFA as phonological, syntactic or stylistic in nature, BCS ALFA is a definite case of a “higher level” sociolinguistic variable, inasmuch as all variables hold the same truth values, and their distribution is based on social factors (specifically, nationality).

Furthermore, despite the status of BCS ALFA as a phenomenon ripe for sociolinguistic inquiry, it is surprising that no major grammar has recognized its sociolinguistic implications or given a detailed (or satisfactory) explanation of the historical origin of these three endings, their subsequent development across different synchronic slices of BCS, or the motivation for their haphazard distribution in contemporary BCS. For the most part, ALFA in BCS has been generally overlooked in the traditional grammars (and listed simply as optional variants separated by a comma in parentheses next to the prescriptive norm (-om) with scant explanation) until recent treatments in Alexander (2006) and Browne (2006)).

Therefore, this lack of diachronic or synchronic detail in the presentation of BCS ALFA has allowed me to attack the problem from a multidisciplinary approach, incorporating aspects of historical linguistics, corpus linguistics, and sociolinguistics.

As discussed above, there have been major criticisms proferred in the past that morphosyntax is inherently not a variationist phenomenon. However, these criticisms only narrow down the possibility of studying purported morphosyntactic variables from a sociolinguistic sense, inasmuch as the main criticism is that all variables should share exact truth values. When such conditions are met, then these criticisms are no longer applicable.

One additional criticism that potentially blocks the way for this study involves the supposed methodological shortcomings of doing a sociolinguistic inquiry on
morphological or syntactic phenomena. The two major shortcomings are: 1) the potential for studies of less frequently occurring linguistic phenomena to be less statistically representative of a group of speakers, and 2) the idea that morphology and syntax are more “inviting” to new elements due to their “leaky” natures, unlike phonology, which has a sharply limited set of members.

2.2 Statistical Representativeness and "Leaky" Systems

This criticism, largely promoted by Milroy and Gordon (2003), involves the concept of the “leakiness” of systems, in this case of higher-level systems, i.e. morphology and syntax, and the potential for social indexing. First, on general frequency issues in higher level and lower level phenomena, Milroy and Gordon state (170):

> Since speakers make use of a sharply limited and therefore frequently recurring inventory of phonological contrasts, realizations of any given variable are likely to show up quite frequently in even a short sample of speech. This is not the case for morphological and more especially higher level syntactic variables.

While Milroy and Gordon are not saying it is impossible that morphological or syntactic variables can bear social indices, their emphasis that phonology (unlike higher level variables) has a sharply limited number of phonetic realizations for any given phoneme marginalizes morphology and syntax. In a certain sense, morphological and syntactic realizations are limited as well. Where phonemes can vary in the way they are phonetically realized idiolectally, the semantics of a group of phonological units (a word) remain unambiguous, i.e. the meaning is still clear although a speaker may have to accommodate to his/her interlocutor’s idiolect to
understand a variable pronunciation (i.e. a sort of lowest common denominator process). On the other hand, any modification of morphological, lexical, or syntactic forms typically results in a modification of meaning (e.g. whatever meaning the affix, lexeme, or syntagma contributed), and this of course can lead to intelligibility issues as such variants arise in a dialogue between speakers of different dialects. Moreover, a speaker’s use of a lexical, morphological, or syntactic variable unfamiliar to his/her interlocutor can even cause a “jarring” effect, so to speak.

One particularly illustrative example of morph-syntactic variation that can cause a jarring effect is the expression need + past participle, e.g. The car needs washed, considered an “Appalachianism”, according to Wolfram and Christian (1976: 96). For someone who speaks a dialect of American English, which has The car needs to be fixed and likely also The car needs fixing, he/she will inevitably be puzzled when hearing needs fixed for the first time. Moreover, once the speaker understands the meaning, he/she will likely attach some indexical information to the expression, such as where the speaker comes from, the speaker’s education level, etc. For many speakers of the Appalachian dialect, both The car needs to be fixed and The car needs fixed are acceptable and do not change the truth values of the utterance. This example would appear to be an exemplification of what should be considered a sociolinguistic variable, inasmuch as social identity is the most important factor in its usage.

However, let us turn to the second “short-coming” of doing higher level sociolinguistic inquiries, according to Milroy and Gordon, which brings into question
the statistical reliability of empirical studies on morphology and syntax due to their so-called “leaky natures” (Milroy & Gordon 2003: 170).

...a sufficient number of tokens of a given type of construction cannot usually be guaranteed to appear in a piece of spontaneous discourse. This difficulty is partly a consequence of the non-finite or “leaky” nature of syntactic systems, which in turn is associated with the susceptibility of syntactic [and morphological—JP] choices to pragmatic and semantic constraints.

First of all, it seems Milroy and Gordon believe that the best way to observe variation is through spontaneous speech. I agree that this is the best way to address phonetic issues (through audio recording or phonetically tagged corpora of audio recordings). However, for morphosyntactic variables, phenomena that are less likely to be captured frequently in spoken discourse, a properly formulated questionnaire that includes many tokens of such variation can address the issue with empirically sound results (Labov 1969; see Chapter 4 of this study). Secondly, much ink has been spilled over exactly how many speakers and tokens are needed in empirical studies of the sociolinguistic variable.

Chesire (1999) and Hudson (1996) view syntactic variables as more difficult to study due to their relatively lower frequency in comparison with phonology. It is logical to view the limited set of units speakers have in their phonological inventories as occurring more frequently than the vastly greater number of morphological units that are realized through this very phonology. Subsequently, Chesire and Hudson have attached a whole series of generalizations about low frequency as supposedly tied to low social indexing. Conversely, they see higher frequency tied proportionately to higher social indexing. Nevertheless, even though the low-high parallelism shared in

12 Milroy and Gordon use “leaky” here to mean “more inviting to new members”.

16
both Chesire’s and Hudson’s views is appealing, neither author gives explicit examples to support this claim, and therefore they fall short of being convincing due to a lack of empirical support. Moreover, I believe Chesire and Hudson are comparing apples (phonology) to oranges (morphology, syntax, etc.), and their view needs to factor in that sociolinguistic “saliency” is tied to both ends of the frequency spectrum.

It has been demonstrated that variables occurring less frequently in language are treated as simply “different” in some way, and this labeling at times entails speaker assessment of the background of a speaker who used such a form. For example, in American English, for many speakers, the distribution of pop and soda, or lightning bug and firefly, for instance, is an “us vs. them” affair. Speakers who use one form exclusively are rarely able, when asked, to correctly label geographical areas of the United States where they believe the other form prevails (Preston & Long 2002).

Generally speaking, however, speaker perceptions of other dialects in the United States are consistently stereotypical (and many times stigmatized), as if all Southerners have their “twang” and New Englanders all sound like John F. Kennedy (Preston 1996/1999).

Even more specifically, in jargonistic speech, speakers can choose to exercise a sort of language “discrimination” by using a less common word instead of a more widely used variant. For example, in a recent (2009) report published by the British House of Commons Public Administration Selection Committee (PASC),13 the members recognize that:

---

13 Available online at: [http://www.publications.parliament.uk/pa/cm200910/cmselect/cmpubadm/17/17.pdf](http://www.publications.parliament.uk/pa/cm200910/cmselect/cmpubadm/17/17.pdf)
politicians and public servants should use language that people find clear, accurate and understandable. We undertook this inquiry because we were concerned that too often official language distorts or confuses meaning. This is damaging because it can prevent public understanding of policies and their consequences, and can also deter people from getting access to public services and benefits… Bad official language deserves to be mocked, but it also needs to be taken seriously.

The committee members understand the sociological divide that can be caused when one group specifically chooses to use language that is incomprehensible to another (in this instance, politicians vs. laymen). For example one minister complains that, while serving for eight years, she kept a “little book of bullocks”, in which she wrote down absurd examples of political jargon, which she believed did not reflect the “language of the real world” (PASC 6). Furthermore, in an interview with the committee members, the well-known British linguist David Crystal offers his take on jargon (PASC 8):

Every group has its jargon. There is no group on this earth that does not have a jargon. It is when that jargon becomes opaque to the outsider, when the people say, “It is not just enough for us to talk to each other, we have to talk to the outside world” and they forget the demands of the audience, then it gets tricky.

Conversely, if speakers of a particular jargon are aware of the demands of their audience, but choose to use a jargonistic term, they do it to separate themselves from their audience, and thus demonstrate their affiliation with a different group.

This type of jargonistic speech is quite common in talk about sports among Americans. For example, a fan of the National Basketball Association’s (NBA) Cleveland Cavaliers might say in imitation of color commentator and former Cavalier basketball standout Austin Carr: “Get that weak stuff outta here!” when describing an
emphatically blocked shot while watching a game among fellow Cavaliers fans. This expression would certainly endear him to his fellow Cavaliers fans, whereas if (s)he says this to a fairly well-informed Celtics fan, the effect might be confrontational. When used among others, the expression has no impact, and repeated usage might cause frustration to an audience. Examples like these are common.

Therefore, I believe the idea shared by Chesire and Hudson of “less frequent” equaling “less socially indexed” is not at all accurate; in fact, it appears that while for phonology this is true, for morphology and syntax the opposite is true. However, Chesire’s and Hudson’s stance is shared by Milroy and Gordon in their discussion of the concept of “sharp stratification” and its implications for sociolinguistic inquiry (Milroy and Gordon 2003:170). They view syntactic variables as “less available than phonological variables as resources for social evaluation and so less likely to be socially indexed” (ibid). Therefore, in this view it would seem that the more frequently a feature occurs, the more potential it has for social indexing. However, Milroy and Gordon seem to contradict this central assumption within the same paragraph: “even relatively frequently occurring morphological variables…do not appear to fulfill the same sociolinguistic function as phonological variables” (Milroy & Gordon 2003:171). In such a view, how can relatively frequently occurring variables of any type, which are frequently available for social evaluation, not be socially indexed?

Phonology seems to occupy a special place for social indexing in Milroy and Gordon’s view. They consider this type of sharp stratification as an “all or nothing” affair (Milroy & Gordon 2003: 171), as morphological and syntactic variables show many fewer gradient patterns of usage, unlike *h-dropping* in Bradford and Norwich
English where the variation observed in the two cities is dependent on phonetic environment. (Chambers & Trudgill 1998: 59). Nonetheless, one could argue that morphology and syntax being “less gradient” in nature could be a sufficient reason to place a high potential for social indexing on them, inasmuch as this low gradience should guarantee that any deviation from pre-existing morphosyntactic patterns would result in salience for the new pattern. This salience would presumably catch the ear of a speaker who lacks the innovative pattern.

The above examples of higher level variation (e.g. need fixed and jargonistic speech) call into question the notion that phonology is more socially indexical because of frequency values. In fact, it really is not logical to even compare phonological systems with morphological systems in terms of frequency. While it is true that there are only a sharply limited set of “sounds” in a language, because these sounds occur so frequently, one would expect any social indexing to occur over a gradual period of time, inasmuch as many speakers can be blind to sound change in progress. For example, a Californian who hears a Bostonian say I parked my car in Harvard Yard, which exhibits the stereotypical Bostonian “r-lessness”, would not be necessarily jarred into incomprehension, so to speak, since the phonological shape of the words remains intact for the most part. As the conversation continues, the Bostonian may use his r-less forms several more times before the Californian sees the full range of usage. Subsequently, the Californian can adopt a pronunciation habit, as it were, and attach social indexing accordingly.

Therefore, stated more formally, if one thinks of higher-level variation and lower-level variation from the perspective of frequency, then it follows that they
occupy two different extremes of a frequency spectrum. However, for frequently occurring elements on this spectrum, critical mass for sociolinguistic saliency is intrinsically high as well, since a low critical mass would suggest that only a few occurrences of a sound can potentially lead to social indexing (which is unlikely). Similarly, for higher level phenomena, their low frequency values guarantee a low critical mass for saliency, since a high critical mass would suggest that higher-level variables do not have the jarring effect as discussed above (which we know not to be true).

On the whole, the above views about social indexing and frequency, the leakiness of morphological and syntactic systems, and sharp stratification, do not give morphology and syntax a “fair shake”, so to speak, from the variationist perspective. Winford (1996) is justified in bringing in the pragmatic notion L-meaning (literal meaning), as introduced by Geis (1994) as the most important criterion in defining the sociolinguistic variable. The fact that all three BCS ALFA endings, -om, -ome, and -omu, share the same L-meaning hopefully will help to dispel the notion that phonology is somehow “more special” in a sociolinguistic sense. Morphological variables sharing the same L-meaning should not be a stable situation if we agree with Lavandera’s notion that it should not be possible for a language to have two ways of saying the exact same thing. This is where the probabilistic, or stochastic (usage) elements of language come into play.

In addition to the above criticisms leveled against studies of morphosyntactic variation, Frederick Newmeyer’s recent contributions to Language (78.4.682-707, 2003/81.1.229-36, 2005) demonstrate the most extreme criticisms of grammatical
variation and its manifestation in speech communities. In his 2003 article, he resoundingly states that “there is no way that one can draw conclusions about the grammar of an individual from usage facts about communities.” He then invokes Lavandera’s work: “Lavandera [1978] pointed out that the choice of phonological variants is purely a social matter. But syntactic variants often differ in the meaning that they convey” (Newmeyer 2003: 696).

In response to Newmeyer’s articles, Gregory Guy (Language 81.3:561-563, 2005) holds Newmeyer accountable for making “little reference to the extensive work in variationist sociolinguistics addressing [Newmeyer’s criticisms].” Guy is fundamentally opposed to Newmeyer’s view of the relationship between a speaker and a speech community, reminding us that the “relationship between the community is an explicit focus in sociolinguistics, and the principal finding is that individuals are grammatically similar as a function of social proximity” (ibid: 562). Furthermore, Guy opposes Newmeyer’s view that phonological variants are inherently different from morphosyntactic ones: “[Newmeyer] goes into the hoary argument that perhaps syntax is different from phonology, but massive work on syntactic variation reveals nonepiphenomenal ‘stochastic generalizations’ as the OCP [obligatory contour principle] case” (ibid).¹⁴ Interestingly, while the OCP has many categorical effects in language, its instantiations can either be categorical or probabilistic.

Riding on Guy’s coattails, so to speak, let us now return to the subject of BCS adjectival long-form variation. While the morphological components of BCS ALFA

¹⁴ Guy is referring to a theory which was born out of the Labovian Variable Rule Theory (Labov 1969), an attempt to account for probabilistic elements in grammar. Newmeyer’s articles stand in opposition to stochastic grammar. OCP is an acronym for the ‘obligatory contour principle’ of phonology, which states that adjacent, phonologically identical elements are disfavored in a string of segments.
are undoubtedly categorical, their distribution appears to be probabilistic. This can be explained in large part by considering the exceptional circumstances in which the Bosnian, Croatian, and Serbian grammars have evolved and considering the effects of dialect convergence between Eastern Štokavian (roughly equivalent to modern-day Serbian) and Western Štokavian (roughly equivalent to modern-day Croatian) speakers.

Serbia and Bosnia endured more than five hundred years of Ottoman rule, while Croatia endured the burden of Roman (Latin), German, and Hungarian language policy. With many Eastern Štokavian speakers fleeing the Ottoman Empire, seeking refuge with their Christian brethren, either in Croatia, Austria, or Hungary, a situation arose where Slavic ethnic groups speaking essentially the same language (Central South Slavic), albeit with dialectal differences, came into contact with one another (Ivić 1972: 78-82). One of these differences would have been the realization of ALFAs in the convergent dialects. I believe such convergence is responsible for the current BCS ALFA distribution, and this is discussed further in Chapter 6, an examination of the diachronic aspect of BCS ALFA variation.

In conclusion, despite the criticisms of Romaine (1984), Milroy and Gordon (2003), and more recently Newmeyer (2003, 2005), that morphosyntax is harder to do from the sociolinguistic perspective, there have been many successful works carried out on AAVE morphosyntactic variation, including studies like Labov and Wolfram’s (1969), Rickford et al. (1991), Mufwene (1998), Rickford (1999), Poplack and Tagliamonte (2001), and Walker (2000); more recent contributions on

15 When speaking of historical Serbian, I will refer to it as “Eastern Štokavian” in order to avoid pinning contemporary national labels on a wide variety of ethnic groups, which shared the territory known as modern-day Serbia.
morphosyntactic variation include Tagliamonte & D’Arcy (2009) and Tores Cacoullos & Walker (2009). Works such as these continue to shed light on non-phonological variables that set standard languages apart from other varieties. Moreover, these works give irrefutable credibility to the notion that higher-level linguistic variables can carry important sociological information, although it may be harder to get at the data.\(^{16}\)

In the next chapter, I present a chronological review of relevant literature about BCS ALFA from both native and non-native sources. I begin with Serbian grammars going back to Vuk Karadžić’s *Kleine Serbische Grammatik* (1824), and end with the latest Croatian grammars. It is surprising that more often than not these sources disagree about ALFA usage, and this confusion culminates in the two latest Croatian grammars (Barić 2005/Raguž 2010), which essentially go against the grain of two hundred years of scholarship on BCS ALFA by making claims of an entirely new distribution of the adjectival long forms.

\(^{16}\) For more discussions weighing the positives and negatives of doing variationist sociolinguistics on higher level phenomena, see Macauley (2002) and Buchstaller (2009), who see the importance of such studies, but call for more standardization of the quantification and reporting methods of sociolinguistics
CHAPTER 3

HISTORIOGRAPHY OF BCS ADJECTIVAL LONG-FORM ALLOMORPHY (ALFA)

Generally speaking, the esoteric problem of inflectional variation in the adjectival declension of Bosnian-Croatian-Serbian has been only superficially discussed, or even many times ignored, until recent treatments in modern grammars (Alexander 2006, Browne 2006). In most modern grammars of BCS, three distinct desinences are listed for the DL adjectival paradigms (BCS ALFA).\(^{17}\) This type of variation is quite rare in the Slavic languages, which generally show contextually based allomorphy, whether the conditioning be semantic (animacy, for example), phonological, syntactic, etc. In fact, no other standard grammar for ANY of the Slavic languages cites parenthetical forms standing alongside normative forms as prescriptively acceptable (this, of course, excludes lexical doublets, which are typically parenthetically cited in grammars).

Moreover, when variation exists in language, a linguist’s job is not only to describe the phenomenon, but also to explain its genesis and rules for usage, where possible. However, many of the early “core” grammars of BCS have confused the issue of BCS ALFA distribution thoroughly by parenthetically citing forms in

\(^{17}\) The ACC/GEN ALFA in –oga is used in a similar manner as the DL ALFAs. Equal attention will be given to AG ALFA below in chapters on historical genesis, corpus analysis, and sociolinguistic fieldwork on ALFAs in BCS.
paradigm charts without explaining, first of all, their intent when using parentheses, and secondly, the meaning of the order in which the long-form adjectives are listed in grammars and handbooks.

There is no universal linguistic principle for using parentheses and ordering forms. However, the OED lists variable grammatical morphemes in their historical order. Unfortunately, BCS grammars do not follow this tradition, and this is clearly demonstrated in the paradigm charts presented in this literature review. The wide variation in ALFA presentation is evidence that the historical ordering never was a tradition in descriptions of BCS ALFA. This leaves the linguist guessing, essentially, whether the *parenthetical forms* should be equated with “less common/acceptable” or “dialectal/colloquial”. Perhaps we should assume the *ordering* of forms as entailing “the form to the left is more frequent than any form occurring to its right”. However, this “guesswork” is based on reasonably foggy notions, as the leftmost form can either mean 1) what grammarians wish to promote as the standard (for one reason or another), or 2) what speakers actually say. Therefore, the ordering of forms presented in grammars may reflect the norms of the dialect of the compiler of a grammar, or his “Academy”, and these norms often reflect intuition-based judgments made by native speaker grammarians about frequency. (For a good example of an ad hoc intuition-based rule, see Barić 2005, Raguž 2010). Moreover, the ordering of forms says nothing directly about the frequency of use for native speakers of a given language when comparing their supradialect (standard language) with its subdialects.\(^\text{18}\) Hence,

---

\(^\text{18}\) This "register vacillation" has typically has been referred to as "style" in modern BCS handbooks (cf. Browne 2004[6], Alexander 2006); a problem which will be addressed at the end of this chapter.
without empirical data to back up these claims, these claims cannot be taken at face value.\(^{19}\)

In general, there seems to be a tendency by BCS grammars to “piggy-back” on prior grammars of the language, at times neglecting language variation and change. Often, grammars will include one norm as the paradigm form to the exclusion of secondary forms. Furthermore, the ordering of forms has changed over time in the presentation of BCS grammars, and this change has not been discussed in the literature despite the problems it poses for students of the language and linguists who refer to grammars from different decades (or even those from the same decade which show drastic differences, e.g. Alexander 2006 and Raguž 2010), and the significance it may hold from the perspective of language change.\(^{20}\) Therefore, there are two questions I ask the reader to keep in mind (although they are not directly answered) when surveying the literature review on BCS ALFA below: (1) Could the haphazard representation of ALFAs in BCS grammars have compounded the modern-day complexity of its use by speakers (as prescriptive norms are taught in schools)? 2) To what degree is the author’s linguistic identity reflected in his/her description of BCS ALFA variation to the exclusion of other dialects?

\(^{19}\) Cf. Chapters 4 and 5 for empirical investigations of ALFA distribution in BCS.

\(^{20}\) Examples of how the ordering of forms varies from handbook-to-handbook will be given below.
3.1. Serbian Grammars

3.1.1 Karadžić, Vuk Stefanović. (1824). *Kleine Serbische Grammatik*

Any historical or synchronic investigation of the so-called “Modern BCS” language must begin with Karadžić’s 1824 *Kleine Serbische Grammatik* (first published under the title *Pismenica srpskoga jezika* (1814), but first made widely accessible through a German translation in 1824), simply because this was the first codification of Serbian as a literary language based on how the “folk” spoke. Moreover, Vuk’s description of Serbian morphology holds up exceptionally well against modern descriptions of BCS morphology, making it a priceless resource for historical and synchronic linguists alike, while at the same time bringing into question how much modern grammars actually reflect the grammar speakers use.

Thankfully, in documenting and representing folk speech in his grammar, Karadžić was careful enough to include not only the typical paradigm charts, but also elements of the grammar which showed variation. For example, for the verb *sipati* ‘to rain/pour’, he gives two possible present-tense stems, *sipa-* and *siplj-* (Karadžić 1824: 75).\(^{21}\) Regarding BCS ALFA variation, Karadžić (1824) was the first to describe the problem. Beginning with *indefinite/abstract* descriptive adjectives, Karadžić gives the following forms for *žut* ‘yellow’ (Karadžić 1824: 41):\(^{22}\)

---

\(^{21}\) This possibly reflects different verbal stems made with this root (cf. Russian *maxat*—*mašet-maxajet*, ‘to wave (3rd sg.’).

\(^{22}\) The accusative is syncretic with the genitive for animate masculine or neuter nouns. Hereafter I refer to these forms as genitive only, which subsumes the accusative.
For possessive and demonstrative adjectival declensions he lists the possible forms for *møj*, which is representative of all possessive and demonstrative adjectives (Karadžić 1824: 59):

<table>
<thead>
<tr>
<th>Gen/Acc</th>
<th>žuta, -oga, -og</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dat</td>
<td>žutu, -ome, -omu, -om</td>
</tr>
</tbody>
</table>

Table 3.1: ALFA for Descriptive Adjectives in Karadžić (1824)

It is difficult to understand the ordering of forms, and Karadžić gives no explanation of them himself. Apparently, for possessive adjectives, longer forms precede shorter forms from left to right. However, other than mentioning the existence of the forms, Karadžić does not explain in which contexts the variable forms are used. Karadžić’s *Grammatik* would begin a trend of over a century of BCS grammars which offer little or no explanation for ALFA variation.

3.1.2 Hraste, Mate and Sretan Ţivković. (1949). *Gramatika hrvatskoga ili srpskoga jezika.*

Hraste & Ţivković mark the first grammar to attempt to explain ALFA variation, looking at its syntactic distribution. They conclude that the long(er) forms

---

23 The variants –*omu* ~ –*emu* are dictated by stem final hard or soft consonant, respectively.
are used as markers of definiteness when an adjective stands without a noun to modify (which is integrated into the adjective as a marker of definiteness) (Hraste & Živković 96):

(a) Govorim o mladome. ‘(I) speak about the young one’.

(b) Mudromu je i jedno oko dosta. ‘For a wise man, one eye will suffice.’

In addition, when a noun is modified by two or more preceding adjectives, the first will usually decline as a long(er) form, whereas the others remain short (ibid):

(c) … po starome skliskom putu… “… along the old, slippery path…”

(d) … jednoga ljetnog dana… “... of one summer day…”

These are two particularly useful statements about ALFA distribution that can be tested through corpus-based analysis and sociolinguistic survey. Moreover, these statements can be interpreted as possibly, though not necessarily, conflicting with the typical view that long/short forms should be equated with “definite/indefinite” since examples (a) and (c) are ambiguous with regard to their definiteness. On one hand, it appears examples (c) and (d) demonstrate the treatment of ALFAs as phrasal “clitics”, in the sense that the addition of G –a and DL –u or –e cannot be added to more than one adjective in a phrase (unfortunately, this possibility is not addressed in their work). On the other hand, it is particularly puzzling why combinations of ALFA + NOUN + ALFA are considered grammatically correct. For example, po starome putu skliskome is considered acceptable. In this context it appears that the optionality of long forms begins with the leftmost adjective, but this optionality can be “renewed” if

---

24 A sociolinguistic questionnaire and corpus-based analysis are the subjects of Chapters 4 & 5 of this dissertation, respectively.
the second adjective is preceded by a noun. Moreover, it is curious why the long forms can occur before a noun only when an adjective follows the noun, and not when there are simple ADJ + NOUN combinations. Oddly enough, although as useful and important as their grammar may be, reference to Hraste & Živković’s work on ALFAs remained absent from English Grammars of BCS until Browne (2006), who slightly revises Hraste & Živković, and Alexander (2006), who has written the most comprehensive survey of ALFAs to date (both who are discussed later in this review).

3.1.3 Belić, Aleksandar. (1962). Istorija srpskohrvatskog jezika / Predavanja Aleksandra Belića

Belić’s 1962 seminal treatise on the history of BCS would follow in Hraste & Živković’s footsteps by further updating the problem of ALFAs in BCS. Specifically, for genitive ALFA forms, Belić claims G –og occurred more frequently than –oga, and that –oga had acquired a more “archaic” status (Belić 141). Similarly, DL –om, -ome were used more frequently than –omu, which interestingly was still used frequently in the dative case, but had become “archaic” in the locative. It is surprising to see, then, that Belić’s paradigm charts are listed as follows:

25 A similar principle seems to explain coordinated participles in OCS: “In a series of coordinated substantivized participles (rarely other adjectives also) the long form usually occurs for the first only: (L 6:47, 49) slyšči slovesa moja i tvorq ja...slyšavqy i ne tvorqy ‘he who hears my words and acts...he who has been hearing and does not act” (Lunt 2001:143). One conceivable hypothesis is that this principle spread to the general adjectival class in West South Slavic, although this is at best sheer speculation without support from empirical data.

26 Conceivably, there could be prosodic factors that led to ALFA constraints, and this topic will be explored below in a subsequent chapter.

27 Belić plays a rather large role in the discussion of the historical genesis of ALFA variation (see Chapter 6).

28 If we invoke Kuryłowicz’s 4th Law of Analogy here (which states that, when an innovating form arises resulting in variation with an older form, the older form acquires a secondary function, and the
This once again brings up the question of “ordering of forms” and its significance. Belić himself describes the leftmost forms as the most archaic, yet he lists G –oga before –og and D -omu before the more common –om and –ome, but L -omu in parentheses presumably to indicate its rare usage. Although it is understandable, in the absence of clear conventions, that an author may choose to cite the forms in his own way, for language learners and linguists, however, this inconsistency can be ultimately frustrating.

### 3.1.4 Stevanović, Mihailo (1964). *Savremeni srpskohrvatski jezik*

Stevanović presents ALFA forms as “takozvani pokretni samoglasnici” (so-called moveable vowels). Morphologically speaking, he considers ALFA forms as being secondary to forms that end in –om and –og. However, he does admit that “ovi se oblici [pokretni samoglasnici], kako je poznato, upotrebljavaju i c njima i bez njih”
(“these forms [the moveable vowels], as is well-known, speakers can use or omit”).

He gives the following parallel examples (Stevanović 262):

(a) DLI plural: *upuštite se zelenim travnjacima* alongside *upuštite se zelenima*\(^{30}\) *travnjacima* ‘Go to green pastures’

(b) G *kuda cveće iz zelenog travnjaka* alongside *kuda cveće iz zelenoga travnjaka* ‘where to put the flowers from the green pastures’

(c) DL *upuštite se zelenom travnjaku* alongside *upuštite se zelenome travnjaku* ‘Go off in the green pasture’

These are important data, as, according to Stevanović, despite being secondary to adjectival forms ending in consonants (e.g. G –*og*, DL -*om*), the ALFA forms can be heard all the same. However, with strings of adjectives preceding a noun, Stevanović states that “jedan [pridev] s oblikom bez pokretnog samoglasnica, a drugi s njim, treći opet bez njega…” (“one [adjective] with a form without a moveable vowel, and the other with it, the third again without it…”) Oddly, however, he cites the following examples, which do not illustrate his rule, but illustrate the possibility of having the first adjective in a string of three with an ALFA, the second without, and the third with (d), or having the first two in a row with ALFAs and the last without (e) (Stevanović 263):

\(^{30}\) This is the DLI plural ALFA in –*a*. However, since Stevanović’s work, this form is extremely rarely encountered in the spoken language.
With regard to the ALFA form –*omu*, Stevanović considers its use old-fashioned, but acknowledges it can still be heard in the western regions of Yugoslavia (essentially Croatia) (Stevanović 264). As for ALFA –*ome*, Stevanović considers it the more central and eastern standard (Stevanović 264).

Needless to say, of course unintentionally, Stevanović makes the ALFA situation even more complicated. The idea that only two out of three adjectives preceding a noun can add an ALFA is one more aspect to attribute to the confusion of ALFA usage, in general. However, the idea that Croats use –*omu* more frequently would later be taken up by Alexander (2006), and subsequently followed up in my Chapter 4, the results of sociolinguistic fieldwork I conducted on ALFA usage in Serbia and Croatia.

3.2 Non-native Grammars of Serbo-Croatian/BCS


Širokova and Gudkov’s *Slavjanskie jazyki* was considered the standard Soviet survey of the West and South Slavic languages. Young Soviet scholars and language students likely would have consulted this book before others as questions about BCS grammar arose. Gudkov’s survey of Serbo-Croatian includes only two pages on
adjectives, but it does present an interesting take on ALFAs. Gudkov refers to the long form adjectives as “relikty” (relics) of the ancient Slavic category of indefinite/definite adjectives. Note that he lists the forms –og(a), -om(e) as only possible in the “polnaja” (full) form paradigm of adjectives, but he chooses to omit the possessive and demonstrative adjectival paradigms (mojega vs. moga for example), which intrinsically cannot can be used in the indefinite/definite sense (e.g. *I saw a/the my dog for Ja sam video mojeg(a) psa), inasmuch as the possessive pronoun is “doing the work” of definiteness, so to speak, making the addition of an ALFA tautologically definite.

Unfortunately, the omission of possessive and demonstrative paradigms depreciates the value of Gudkov’s survey of ALFAs. Moreover, he only lists the pronominal adjectives in their nominal forms as “pritjažatel’nye” (possessive), ignoring the somewhat chaotic morphology of their paradigm. ALFA variation is much more puzzling for the possessive adjectives since they supposedly lack long forms (in the semantic sense of definite/indefinite), but still use so-called “long-form endings” to create G mojega/mojeg/mog/moga and DL mojemu, mojem, mom, mome, for instance. Interestingly, Gudkov omits the optional vowels D –u and IP -a from his paradigm chart altogether, and he does not mention their occurrence in his discussion of adjectives.
3.2.2 Browne, Wayles and Theresa Alt. (2004). *BCS (Bosnian, Croatian, Serbian)*

**Grammar**

Browne and Alt’s *BCS Grammar* is included in this literature review for two reasons: (1) it is the most accessible BCS grammar currently available thanks to its free availability on the internet, (2) it was written in large part by Wayles Browne, one of the world’s leading scholars on BCS. In their section on adjectives, Browne and Alt list the long and short\(^{31}\) adjectival forms of *nov* ‘new’:

<table>
<thead>
<tr>
<th>Long</th>
<th>Short</th>
</tr>
</thead>
<tbody>
<tr>
<td>G novog(a)</td>
<td>G novog(a), nova</td>
</tr>
<tr>
<td>DL novom(e,u)</td>
<td>DL novom(e,u), novu</td>
</tr>
</tbody>
</table>

---

**Figure 4.7** Browne & Alt’s (2004) long and short adjectival paradigms

The only visible difference from prior grammars presentations of ALFAs are the addition of the short forms *nova*, *novu*, which Browne and Alt claim are more frequently heard in Croatia, but the optional long forms for the genitive –*a* and DL –*e*, –*u* are basically left unexplained. Notice that, in contrast to Gudkov’s grammar, *novog* and *novom* qualify as short forms in Browne and Alt’s grammar, and DL –*u* and Ipl. –*i*

---

\(^{31}\) Short forms *nova* and *novu* do not exhibit variation in BCS, nor do they change the truth conditions when replacing a long form. However, Alexander (2006: 52) believes they are restricted to certain set phrases.
are included as optional vowels. However, the authors do not explain in what contexts this “optionality” takes place.

Browne and Alt also explain the contrast of short and long forms in the nominative (e.g. \textit{nov} vs. \textit{novi}) as being semantically motivated, and they add, “Since Vuk Karadžić they have been explained as answering the questions \textit{kakav?} ‘of what sort?’ and \textit{koji?} ‘which one?’ respectively” (Browne & Alt 2004: 38). However, they do not say whether this is true of the oblique case ALFAs. Browne and Alt do reveal a morphological distribution of \textit{DL –e} and \textit{–u} according to stem-type: “Soft stems differ from hard only in nominative-accusative neuter singular long \textit{lošē}, short \textit{loše} ‘bad’, masculine-neuter genitive \textit{lošeg}(-a), dative-locative \textit{lošem}(-u)” (Browne & Alt 38). This claim is verified by a check of the Croatian National Repository and the Oslo Corpus of Bosnian Texts (both available online), which return zero hits when searching for the DL of soft-stems in \textit{*–em(e)}.

Despite this descriptive contribution to the general literature on BCS ALFAs, the distribution of \textit{–e} and \textit{–u} as optional vowels for hard-stems (the vast majority of adjectival stems) is left a mystery. One might wonder whether Browne and Alt’s ordering of optional vowels in the DL is not their informed intuition about the language or the same “piggy-backing” that began with Vuk’s \textit{Grammatik} in 1824, and which has lacked the proper empirical study to finally resolve the issue. This is not a

---

32 This raises another problem that is typical of discussions of BCS ALFA—getting the terminology right. Despite the confusion over how to describe these forms, throughout this dissertation I use ALFA to refer to long G and DL base forms in \textit{–og} and \textit{–om} and their “longer” counterparts \textit{–oga}, and \textit{–ome}, \textit{–omu}, respectively.

33 Soft-stem adjectives end in a palatal consonant. An initial /o/ of a suffixal morpheme will be realized as /e/ via fronting assimilation.

34 For the most part, vowel length does not play a role in this dissertation. However, where relevant, vowel length is indicated.
criticism of Browne and Alt, but simply a reminder that the optional vowels occur so rarely that it is hard to say just when and how they are used without an empirical study.


In comparison to Browne and Alt’s 2004 Grammar, the core section here on adjectives remains mostly unaltered except for one key addition: “the movable vowels” (a), (e), (u) tend somewhat to appear in phrase-final position, otherwise not: o tome… ‘about that’, ‘o tom psu’… ‘about that dog’…” (Browne 2006: 323).

This is a critical new wrinkle absent from prior English grammars of BCS. Nevertheless, this appears to be a less informative restatement of Hraste and Živković (see above), since Browne omits the constraint on long forms when adjectives precede a noun, an important notion inasmuch as it is tied to the category of definite/indefinite. The ordering of forms remains the same as Browne and Alt (2004).

3.2.4 Alexander, Ronelle. (2006). Bosnian, Croatian, Serbian: Grammar with Sociolinguistic Commentary

Ronelle Alexander has done an exceptional job of combining previous scholarship on BCS ALFAs, and she follows up on Stevanović’s (1964:263) notion that DL –omu is “more Croatian”. She divides the descriptive adjective paradigm

---

35 Moj and tvoj, for example, lack short forms, although they do have contracted forms of full forms, e.g. G mog ← mojega, DL mom ← mojem. However, the contracted forms are not used like other short forms (i.e. they are not used to mark definiteness. Alexander also notes that Bosnian and Serbian only use the contracted forms, while Croatian uses all possible forms (Alexander 2006: 52).
into three categories: short, long, and longer. In the nominative, short and long can still be used to mark indefinite and definite respectively. However, in the G and DL cases, Alexander states: “the distinction of short vs. long no longer corresponds in all instances to the meaning of indefinite vs. definite; in other words, language change is currently underway” (Alexander 53). She claims that the longer forms are more typically heard in Croatian, and that the longer form is required when standing alone. It is noteworthy that this supposed requirement goes against Hraste & Živković (1949) and Browne (2006), suggesting either that previous accounts factor the Croatian data into their Bosnian and Serbian data instead of considering them separately (thus the confusion over variation), or that Croatian has recently developed this obligatory longer form constraint.

Furthermore, Alexander is the first linguist to suggest (in print, at least) that the category of definiteness in BCS is losing ground to a new sociolinguistically based distribution.36 Alexander states: “Although [attributing the use of ALFAs to definiteness] was once a correct description of the language, it is now true only of the nominative (and not true in all instances about that case either)…it is more correct to say simply that two different forms are possible in the nominative” (ibid). In closing, Alexander leaves the reader with a rather provocative suggestion: “…there is no consistent difference in meaning [among the optional vowels], other than the rather fluid one of style” (ibid) [my emphasis].

Style is a term that has been often used by sociolinguists as a “catch-all” for a number of different intra-speaker variable phenomena, among which are included: 1)
shifts in usage levels between different groups of speakers (dialects), and 2) shifts in registers (formality), dependent on situational factors (Schilling-Estes 2002: 375). In his New York City survey, Labov convincingly demonstrated that style is an independent (extra-linguistic) variable which can affect dependent speech variables (1966: 90-135). As defined by Chambers (1998: 32), *style* is primarily a pragmatic strategy which speakers use to accommodate their speech to sociological variables:

> [Speakers] are capable of considerable discrimination, spontaneously, and almost instantaneously, concerning the degree of familiarity between the participants in a conversation… Observations about speech styles fall squarely into the domain of sociolinguistics… The sociolinguistic relevance comes about because our ability to judge the formality of a conversation comes about by linguistic cues.

Consequently, considering the above discussion of style, Alexander (2006: 52) is suggesting that, while in previous investigations of ALFA linguists have focused on dependent speech variables, such as phonological rules, morphonology, or syntactic operations (in the structuralist tradition), for example, the modern distribution of ALFAs in BCS is dictated by independent variables, such as dialectal differences and register vacillation (the traditional foci of modern sociolinguists). Therefore, Alexander’s description of ALFA is to be praised by finally dismissing notions that what was true of BCS ALFA in older grammars reflects the current situation, consequently ending the tradition of grammar “piggy-backing”. Specifically, Alexander has done much to move the question of ALFAs away from their diminishing correlation with definite vs. indefinite. However, most importantly, she has done a judicious job of explaining ALFA usage in the dialects of BCS.

---

38 In actuality, Fischer (1958) is typically accredited with being the first to consider social factors on linguistic variation.
Notwithstanding these contributions, Alexander’s description of ALFA usage as being stylistically motivated points linguists in a fairly vague direction, inasmuch as any instantiation of the term *style* lacks clear meaning without being bolstered by some examples from real speech. As discussed above, the term *style* can have a variety of meanings, and without an explanation of how exactly ALFA usage varies stylistically, BCS linguists\(^{39}\) are not any closer to knowing in which contexts speakers produce the forms. For example, the idea that ALFA –*omu* sounds more Croatian is not far-fetched, especially coming from linguists with the excellent BCS intuition of Ronelle Alexander and Wayles Browne. Or perhaps there are differing rules of usage in the dialects, but these intuitions need to be backed up by empirical studies, which then would give substance to Alexander’s invocation of style here.

### 3.3 Croatian Grammars

#### 3.3.1 Bartol Kašić. (1604). *Institutiones linguae Illyricae* (Osnove ilirskoga jezika)

Kašić’s *Institutiones* is the oldest extant grammar of the language once titled “Illyrian” (identifiable as the language spoken roughly on the modern-day territory of Croatia). The grammar is quite impressive for its time in its attention to detail. Therefore, it is surprising to notice that for the DL of adjectives Kašić only cites forms in –*omu*, without any mention of variant forms.

\(^{39}\)Outside the scope of this dissertation are the pedagogical issues that arise when variation is presented in a textbook, but left unexplained.
3.3.2 Ardelio della Bella. (1728). *Instruzioni grammaticali della lingua illirica*

Based in large part on Kasić’s 1604 grammar, Ardelio della Bella updates the adjectival paradigms by his inclusion of definite and indefinite forms (for the nominative case only). However, he does not mention any forms for the DL other than –*omu*. 40

3.3.3 Vinko Pacel. (1860). *Slovnica jezika hrvatskoga ili srbskoga*

Just a decade after the Vienna Literary Agreement, Pacel begins his grammar by continuing in the spirit of Illyrianism (a movement for the unity of South Slavs) (Pacel vii.):

> Ne znam koja je bieda, ako nije neznanje poviesti, naše predje obuzela bila, pa i mnoge današnjike obuze te sumnjamo kakovim jezikom govorimo, koje li mu ime; valjda za to, jer mu je i odviše imena: Hrvatski, Slavonski, Dalmatinski, Bosanski, Hercegovački, Srbski, Crnogorski, Raški il Racki, Slovinski, Ilirski, Jugoslavenski.

I don’t know what the problem is, perhaps if it isn’t a misunderstanding of history, our ancestors overcame it, but today we are unsure of which language we are speaking, what to call it; this is probably the reason why there are so many names for it: Croatian, Slavonic, Dalmatian, Bosnian, Herzegovian, Serbian, Montenegrin, Rashkan, Ratskan, Slovenien, Illyrian, Yugoslavan” [my translation—JP].

As tongue in cheek as Pacel’s listing is of names for what he considers “one” language, an important point to take out of his introduction is that he believes in the

---

40 In his discussion of the earliest linguistic isoglosses of the South Slavs after their migrations to the Balkans, Ivić (1972) mentions several well-know features which are divided along West-East lines, such as the collapse vs. the continued distinction of the *jers*, b) the loss vs. the retention of epenthetic *l*, etc. I believe DL syncretism should be included in this discussion. None of the pre-Croato-Serbian grammars (which I present in this historiography) published before the signing of the Vienna Literary Agreement cite DL forms in –*om* and –*ome* could be chalked up to a simple accident of history, I believe it perhaps suggests that –*ome* was likely a peripheral dialectal feature in the Croatian territories where Serbs resettled during the Turkish occupation. This suggests that perhaps DL syncretism was much cleaner in Croatian dialects than in Eastern Štokavian.

42
unity of a number of dialects, and his grammar is a reflection of what he considered to be the most common elements of all these dialects. In Karadžić’s 1824 grammar, we have already seen the introduction of all three DL forms –om, -ome, –omu, G –og and –oga, and DLI plural –im and –ima. Judging by Pacel’s philosophy of the linguistic unity of the dialects enumerated above it would follow then that his presentation of his language would include Serbian elements if they were part of his own speech. Pacel’s grammar is the first to list –omu and –om as separate morphemes, the former as the dative masc./ntr. ending and the later as the locative masc./ntr. ending. Still, however, there is no mention of DL –ome. Interestingly, he only cites –oga for the genitive (and not –og), and for the DL plural he cites only –ima (and not –im). It appears that Pacel’s grammar may be the earliest extant evidence of ALFA in Croatian, perhaps even in its formative period.

3.3.4 Maretić, T. (1899). Gramatika i stilistika hrvatskoga ili srbskoga kniževnog jezika

Apparently unsure of how to treat the problem of ALFA variation in Croato-Serbian, Maretić cites Vuk Karadžić’s work (1824) throughout his discussion of the adjectival paradigms (Maretić 7). After listing the following long form adjectives from Vuk’s grammar, njegovoga ‘his-G’, njegovome ‘his-DL’, Petrovome ‘Peter’s-DL’ he states:

Vuk držeci se naroda upotrebljava u svojim djelima i jedne i druge oblike, a opet se dobro zna, da su dulji oblici u ovakvijih pridjeva nerpravilni, a kraći su pravilni…Tako isto piše Vuk ne samo pravilno…premda Vuk jedno i drugo piše, mi valja da odbacimo nepravilan, a zadržimo pravilan oblik.
In his works, Vuk maintains that the folk can use both the long and short forms, but it is well-known that the long forms in such adjectives are incorrect, whereas the short forms are correct. Thus Vuk did not explain this correctly. Despite both forms being possible in Vuk’s work, we should consider this incorrect and uphold the correct form. [my translation]

While Maretić admonishes Vuk for not distinguishing “proper” vs. substandard usage of possessive adjectives, in his discussion of the literary languages of Croatia and Serbia just four pages earlier, Maretić uses long forms (underlined in the following passage), suggesting that what was substandard for possessives was entirely normative for descriptive adjectives (Maretić 3):

Hrvatskom ili srpskom kniževnom jeziku služi u najširem smislu za osnovu štokavski govor, koji se oštro razlikuje od druga tri glavna govora naroda našega: od čakavskog i kajkavskog i torlačkog...bugarskom jeziku...

The štokavian dialect, which is sharply distinguished from the other three other main dialects in the widest sense serves as the basis for the Croatian or Serbian literary languages, čakavian, kajkavian, and torlak...for Bulgarian...[my translation—JP]

Maretić’s Gramatika is the first Croatian grammar to cite forms of the DL in –ome.

The prevalence of forms in –ome could have been caused by the collective pressure of several decades of Serbian Štokavian prescriptive norms on Croatian Štokavian.

Unlike in Pacel’s (1860) grammar, where we see the first signs of ALFA in the Croatian standard, Maretić’s grammar shows that ALFA is a recognized part of the standard language by the turn of the century.
3.3.5 Barić (2005), Raguţ (2010)

In the latest grammars of Croatian, Barić (2005) and Raguţ (2010) both present the same phonetically motivated rule for ALFA usage. Raguţ states the rule as follows (2010: 91):

Produženi nastavci (s pokretnim –a u nastavcima –oga/-ega -a u GAjd m. i sr. roda, zatim u DLI mn svih rodova u oba tipa deklinacije, te navesci –u i –e u DLjd m. i sr. roda zamjeničke deklinacije) nisu obavezni, ali su česti, češći u pisanome jeziku.

The lengthened endings (with the moveable –a in the endings –oga/-ega, -a in the GAsg. masc./ntr., and the DLI of all genders in this declension class; the endings –u and –e in the DL masc./ntr. of the pronominal declension) are not obligatory, but are common, more common in the written language. [my translation—JP]

Raguţ’s use of the terms “non-obligatory” and “more common” sounds like quantifiable claims, but he does not present any empirical basis for these claims. As he does not define what he means by “non-obligatory” and “common”, his understanding of the ALFA distribution ultimately remains an open question.

Moreover, the fact that Raguţ does not present empirical data to support his claims is surprising, inasmuch as the Croatian language has two very good online corpora which can be easily queried to justify his claims about the written language. Furthermore, a comparison with Raguţ’s 1997 grammar shows that his 2010 grammar has chosen not to update his claims despite the availability of these online corpora, as his section on ALFAs is repeated verbatim (Raguţ 2010: 91).

Raguţ then makes his most interesting claim, proferring a phonetic explanation for the distribution of ALFAs (ibid):

U nastavku –og(a)/–eg(a) –a se obično dajaje ako imenica počinje kakvim sličnim suglasnikom: novoga kruga, zatim ako se među više atributa uz
Raguž presents two tendencies for the same phenomenon: a) a phonological rule for the breakup of a G or DL adjective if the following word begins with a similar consonant, a velar (for the G) or a labial (for the DL); and b) one of two or more consecutive adjectives will distinguish itself by adding an ALFA, regardless of whether the following word begins with a velar or labial. Let us take a hypothetical example that can test these tendencies: *Nema dobrog hrvatskog kniževnog priručnika* ‘There isn’t a good handbook of the Croatian literary language.’ According to Raguž’s rules, all adjectival forms could have ALFAs, e.g. *Nema dobroga hrvatskoga kniževnog(a) priručnika*. The first adjective, *dobroga* adds –a, fulfilling tendency a): that the ALFA –a splits –og from a velar consonant that is the first segment of the next adjective; the next adjective *hrvatskoga* adds –a as well, for the same reason; and *kniževnoga* could add –a according to the second tendency, that one of two or more consecutive adjectives distinguishes itself, depending on what Raguž means by “distinguishes itself”. In prior grammars it has been stated on multiple occasions that only one ALFA can occur in a sequence of two or more adjectives (Hraste & Živković 1949, Browne 2006; however, Stevanović (1964) allows for two consecutive ALFAs),
and that one ALFA adjective must be the first in the sequence. Three consecutive
adjectives with ALFAs would certainly be at odds with prior accounts.

In order to verify Raguž’s purported tendencies, corpus analysis can be
carried out to give substance to his use of the terms “non-obligatory” and “more
common”. Conveniently, as Raguž’s grammar is written in his native Croatian, his
preface can handily serve as a small, yet demonstrative corpus. In his preface (1½
pages in length) of the grammar there are many occurrences of ALFAs:

na štokavštinu bosanskoga tipa (ikavskoga govora)...programom hrvatskoga
narodnoga preporoda (ilirskoga pokreta)...funkciju općeg, zajedničkoga jezika
svih Hrvata...na kajkavskome jeziku...na hrvatskome jeziku...iz najnovlji jega
vremena...o zajedničkome životu...između hrvat skoga i srpskoga jezika...

For štokavian of the Bosnian type (ikavian dialect)...the program of the
Croatian folk revival (the Illyrian movement)...the function of a common,
unified language for all Croats...in kajkavian...in Croatian...from the most
recent times...about a united life... between the Croatian and Serbian
languages. [my translation—JP]

I count at least nine cases (underlined above) where Raguž violates his proposed
tendencies in his preface. It is surprising that Barić (2005) adheres to Raguž’s
untenable tendencies, given that she is the leading Croatian grammarian of Croatian,
and her work has been generally well-accepted in the West.

Furthermore, if this is partially a phonological explanation for the occurrence of
ALFAs in Croatian, one would expect this phonological “tendency” to be applicable in
all contexts. Therefore, the phrases momak #kupuje ‘The boy is buying’ and Ja sam#
muškarac ‘I am a man’ should both have the alternate forms momaka #kupuje or
momake(u) #kupuje and Ja sama# muškarac or Ja same(u)# muškarac. Of course,

41Ironically, however, Raguž does not follow his tendencies in his explanation of them: u pisanome
obliku ‘in written form’ (ibid).
speakers do not have such alternate forms, inasmuch as the application of such a phonological tendency would disrupt the morphological integrity of the verbs in both examples. This is an argument that the G –a, and the DL –e –u are strictly morphological elements, given that they possess grammatical meaning, and are not merely epenthetic in nature.

Despite Raguž’s inconsistency in using ALFAs according to the tendencies he presents in his grammar, two insights, one his and one my own, can be taken from the above discussion. His insight is that the claims that –u and –e are assigned to the D and L, respectively, have not been “properly justified” (Raguž 2010: 91). This is an important criticism of previous accounts (Alexander 2006) that assigned –u to the D, and –e to the locative. Moreover, this unjustified belief that one DL ending is preferred by speakers partially explains much of the confusion in grammars of the past two centuries regarding the ordering of forms for DL ALFAs. This leads to the second insight, my own, namely that after examining Raguž’s own usage of –e, I have found that he prefers DL –e over –u when an adjective is governed by a preceding preposition that requires the D or L. The interaction of prepositions and ALFAs plays a significant role in Chapter 5.

Continuing to explore the confusion over ALFA usage in BCS, in the next chapter, I present sociolinguistic fieldwork that I conducted in Bosnia, Serbia, and Croatia in the summer of 2009. I address many of the issues raised in this chapter on supposed tendencies for ALFA usage and the claim that ALFAs are more “Croatian.”
CHAPTER 4

AN ALFA SOCIOLINGUISTIC QUESTIONNAIRE

While ALFAs have been described in the literature in an ever-evolving fashion, seemingly as if each new grammar has added or subtracted some context for their usage, they have more recently been described as “stylistically” marked (see the discussion of Alexander (2006: 53) in Chapter 3). In search of a more rigid definition of what “style” means in the context of ALFAs, with the approval of the IRB, I conducted sociolinguistic fieldwork spanning Bosnia, Croatia, and Serbia, involving the distribution of a questionnaire on ALFA usage. However, before discussing the makeup of the questionnaire, and the subsequent results, let us return briefly to the notion (introduced in Chapter 2) that doing morphosyntax from a sociolinguistic perspective can be quite challenging. While Chapter 2 focused on the controversy over whether morphosyntax can be sociolinguistic in nature, the next section addresses the many problems a sociolinguist faces in the composition and distribution of a questionnaire on morphosyntax.

42 Exemption was received on June 6th, 2009 under Category 2; project number: 2009E0513.
4.1 Quantitative and Qualitative Aspects

When choosing a method of sociolinguistic inquiry, many times the linguist is faced with the predicament of ensuring that the quantitative (statistically valid) and qualitative (linguistically descriptive) sides of the phenomenon under scrutiny are equally addressed. The question of statistical validity is one that has spurred much disagreement in the field. The most critical aspect involved in the dispute is the “weighing of costs of achieving statistical representativeness against the limited additional benefits in might provide” (Milroy and Gordon (2003:26). While statistical representativeness is undoubtedly always a virtue, there are a few cases when some of the leading figures in sociolinguistics, including William Labov and James and Lesley Milroy (cf. also Macaulay 1977) engage in sociolinguistic inquiry without being overly concerned with achieving statistical representativeness, yet these studies turned out to be valuable contributions to the field of sociolinguistics.

In justifying this different standard of statistical representativeness, Labov suggested that investigations of linguistic usage, as a principle, yield less heterogeneous data than other studies which intrinsically allow for a much greater variety of responses, such as surveys of dietary preferences or political polls (Labov 1966: 180-1). Sankoff agrees, offering her own revision of this principle suggested by Labov (Sankoff 1980: 51-2):.

If people within a speech community indeed understand each other with a high degree of efficiency, this tends to place a limit on the extent of possible variation, and imposes a regularity (necessary for effective communication) not found to the same extent in other kinds of social behavior. The literature, as well as our own experience, would suggest that even for quite complex communities samples of more than about 150 individuals tend to be redundant, bringing increasing data-handling problems with diminishing analytical returns. It is crucial, however, that the sample be well-chosen, and
representative of all social subsections about which ones wished to generalize

Below is a closer look at Labov (1966) and Milroy & Milroy (1978), which best illustrate the problem of getting appropriate sample sizes.

4.1.1 Quantitative Aspects of Sociolinguistic Fieldwork: Labov’s Lower East Side Survey (1966) and The Belfast Doorstep Survey (Milroy and Milroy (1978))

Labov’s 1966 Lower East Side survey originally had a sample frame of 340 respondents using telephone directories and the electoral register. However, many factors such as death, change of residency, refusal to cooperate, and native vs. non-native speaker status reduced his final total of respondents to 88. Clearly, Labov’s survey was no longer statistically random. Now, from the smaller group of 88, Labov had to factor in other potential biases, such as the exclusion of speakers whose phone numbers were unlisted or those who were unregistered to vote. However, he settled on this final number, which was much lower than the number originally had in mind. Nevertheless, this is considered to be a seminal work in sociolinguistics. One might of course ask why sociolinguists did not hold Labov to a higher level of statistical scrutiny, but this suggests that Labov’s data were deemed valuable in his field irrespective of their statistical representativeness, given that Labov revealed patterns of variation.

Similarly, James and Lesley Milroy appeared to understand the limited additional benefit that statistical rigor provides while carrying out their survey of five communities of Belfast, Ireland (1975–1981). Using recorded conversations through
participant observation, Milroy and Milroy set a target quota of 16 participants (8 male, 8 female) per community. Having reached their quota for participant observation, the investigators designed a “doorstep” survey intended as a complement to their participant observation data. In the end, a total of only 60 speakers (32 males, 28 females) were selected, 20 fewer than they had gathered for their participant observation investigation.

Both Labov (1966) and Milroy & Milroy (1978) have ultimately been considered valuable contributions to the field of sociolinguistics, despite not being based on large sample sizes of the communities. Therefore, it appears human language, not only in the field of phonology, but also morphosyntax, should be able to be studied from the sociolinguistic perspective, even in the absence of large sample sizes.

However, there exists another hurdle the sociolinguist has to jump over, so to speak, that is choosing the appropriate methods of sociolinguistic inquiry for different situations. Questions such as: “Will speakers be comfortable with tape-recorded interviews?”; or “How can I make filling out a sociolinguistic questionnaire an attractive enterprise for native speakers?” These methodological issues and others are discussed in the following section.

4.1.2. Qualitative Aspects: Optimum Method of Sociolinguistic Inquiry for Less Accessible Languages

Unfortunately, where the Milroys succeeded in gaining the permission of the native speakers of Belfast, Ireland for tape recorded interviews, a similar project for
the Bosnian/Croatian/Serbian languages would be doomed from the start. The most important aspects of any interview include securing the trust of the consultant and guaranteeing the benevolence of the investigator. Inasmuch as many of the speakers of the BCS dialects have experienced a tumultuous political history and hold unfavorable views of the West, guaranteeing confidentiality and benevolence in return for tape-recorded sessions is simply not enough to earn the trust of consultants. However, the recording of oral speech is not the only method of inquiry that has proven successful.

Collecting data through written surveys is an established practice in many other scientific fields (sociology, psychology, etc.). In fact, written questionnaires already have a long history of use in studies of Northern German dialectology (Chambers and Trudgill 1998; Chambers 1998b). Moreover, Chambers (1998a) has demonstrated that data gathered from written surveys are statistically no less reliable than questionnaires administered (orally) by a fieldworker.

4.1.3 The Golden Horseshoe Project

A recent example of success using questionnaires has been demonstrated by the Golden Horseshoe project, directed by Chambers. Chambers gathered invaluable data on Canadian English through the wide distribution of postal questionnaires (Chambers 1994). Sampling from a wide range of demographics, the data gathered from the questionnaires has shed light on linguistic variation and change of Canadian English. A very large amount of labor on the part of the fieldworker would be required for such data to be revealed through oral interviews. A questionnaire can reduce this work load significantly, depending on the linguistic phenomenon under investigation.
Another apparent benefit of distributing questionnaires is the potential removal of the Observer’s Paradox, described by Labov as follows: “the aim of linguistic research in the community must be to find out how people talk when they are not being systematically observed; yet we can only obtain this data by systematic observation” (Labov 1970:32). By “systematically” removing the presence of the linguist from observation, we remove the bias caused by systematic observation. It could be argued that postal questionnaires give speakers too much time to “reflect” on their speech, an undesirable effect for investigations of colloquial speech. However, having the linguist present in a “reduced” role and utilizing misdirection can offset this potential pitfall.

First, before discussing how the Observer’s Paradox played a role in the ALFA questionnaire, let us turn to the composition and administration of the questionnaire.

4.2 The ALFA Questionnaire

The ALFA questionnaire was distributed over a two-week span in Bosnia, Serbia and Croatia during the summer of 2009. With a low likelihood that speakers would agree to be tape-recorded by a foreign non-native speaker of BCS (especially given the recent military conflicts between Bosnia, Croatia, and Serbia and the United States’ role in these)\textsuperscript{43}, the appropriate medium for an investigation of BCS ALFA turned out to be an anonymous questionnaire. Moreover, an anonymous questionnaire maximizes the potential for a high response rate, given that it can be done rather quickly. Not only does it take a high input of labor from the linguist, but speakers are

\textsuperscript{43} For a chronology of developments within the literary languages of BCS, concomitant with the military conflicts of the 1990’s, cf. Greenberg (2004: 161).
also generally more reticent to participate when they know they are being recorded (Observer’s Paradox). Below are the details of the questionnaire.

4.2.1 The Structure of the Questionnaire

The questionnaire comprises fifteen sentences containing a variety of contexts where ALFAs could possibly be used. In addition to including sentences about ALFA usage, demographic information was collected for each respondent to discover or rule out any divisions of ALFA usage of a sociological nature. Speakers also had the option of commenting on where or who might use forms they disallow—issues of perceptual dialectology.

Below is a comprehensive list of criteria structuring the composition of my survey (as approved by the IRB) and demographic information on respondents:

Demographics

(i) Age

Overall, 71 respondents filled out the questionnaire. Due to an IRB restriction regarding research with minors, only speakers who were above the age of eighteen were targeted. Despite this restriction, the questionnaire was distributed randomly to willing participants of all ages above 18, and therefore a wide sampling of age was achieved. Thirty-two respondents between the ages of 18–30 participated, 22 between the ages 31-50, and 16 between the ages of 51–79.

---

44 Copies of the ijekavian and ekavian surveys can be found in Appendix C, along with the English translation.
(ii) Sex

The surveys were distributed with parity of the sexes in mind, although men were much less willing to participate than women. As a result, the workload put in to get the 30 male speakers was much higher than to recruit women. Ultimately, the final total reflects this problem, with 30 men and 41 women.

(iii) City and Country of Birth

Solely native speakers of BCS from the countries of Bosnia, Croatia, Montenegro, and Serbia were targeted in the distribution of this questionnaire.

(iv) City and Country of Residence

Inasmuch as the hostilities of the 1990’s in the Former Yugoslavia led to much cross-migration of BCS speakers, including city and country of birth and residence as important criteria was necessary in order to shed light on possible acquisition of innovative usage vs. the retention of native patterns, speaker to speaker.\textsuperscript{45} However, having only a two-week window to distribute of the questionnaire, I found the bulk of the ekavian respondents in Belgrade, 10 Bosnian ijekavian speakers in Sarajevo, and the rest of the ijekavian respondents from various parts of Croatia, specifically Zagreb, Osijeka, and Split (here I even found a few ikavian speakers).

(v) Ethnicity

Ethnicity was included as an important sociolinguistic category as I interviewed not only Croats from Croatian, but also Serbs living in Croatian (also

\textsuperscript{45} Greenberg (2004: 164-167).
Bosnians living in Serbia). Potential differing ALFA usage between Serbs and Croats living in Croatia could reveal social indexing tendencies (as revealed through respondent commentary) by one group or another. Generally, however, 

*Country of Birth* rarely differed from *Ethnicity* (e.g., born in Croatia, Croatian).

### 4.2.2 The Linguistic Criteria of the Questionnaire

The following environments for ALFA usage were included in my questionnaire, all based on tendencies proffered or disputed by various BCS grammarians in Chapter 3. All sentences from the questionnaire that reflect a specific tendency are included to the right in parentheses for reference (some questions may be referenced for multiple ALFA environments).

(i) Successive ALFAs (Questions 2,3,5,8)

Hraste & Živković (1949) and Browne (2006) claim that when a noun is modified by two or more preceding adjectives, the first will usually decline as a long(er) form, whereas the others remain short.\(^{46}\)

(ii) Definiteness and ALFAs (Questions 7, 12, 15)

Traditionally, ALFAs have been explained as the semantic “definite”, in opposite to short “indefinites”. However, Alexander (2006: 53) has stated that this is no longer the case outside the nominative. Hraste & Živković (1949) and Browne (2006) conclude that the long(er) forms are used when the adjective stands without a noun to modify (which is implied). Therefore, this questionnaire includes sentences G

\(^{46}\)Hraste & Živković (1949: 96)
ALFA –oga and the DL -ome, and –omu for descriptive adjectives occurring without a following noun in an adjectival phrase.

(iii) Stand-alone Pronouns (Questions 6,9,10,11,13,14)

While demonstratives and other pronouns have ALFAs like other adjectives, the distinction between a Croatian preference for DL -u and Serbian for –e has been maintained in the literature (Alexander 2006, Browne 2006). The questionnaire also addresses whether speakers show similar usage patterns for possessive and demonstrative adjectives in comparison with descriptive adjectives.

4.2.3 Native Speaker Assessment of Sentences

Speakers were asked to judge each sentence’s “well-formedness” on a scale of 1–4. ‘4’ is considered well-formed and an exact match to the speaker’s usage. ‘3’ has been heard by the speaker, but (s)he does not prefer such usage. ‘2’ is considered a conceivably well-formed sentence, but the speaker lacks such usage. ‘1’ is considered ill-formed by the speaker and unheard of in his/her dialect.

For responses ‘3’ and ‘2’, speakers were asked to answer questions about “where the usage might be heard”, “who might utter such a sentence” and “what (s)he (the respondent) would say instead”. 47

4.3. Removing the Observer’s Paradox from My Sociolinguistic Survey

47 Speakers were not asked to respond for sentences which they rated ‘1’, inasmuch as ‘1’ signified that the sentence was totally ungrammatical for a given respondent.
Returning to Labov’s notion that subjects are less likely to give authentic responses when in the presence of the observer, I paid careful attention to the construction of my questionnaire and its distribution with the ultimate goal of removing the Observer’s Paradox.

4.3.1 The Use of Misdirection

As the reader is well aware by now, the foremost BCS linguists have stated that the ALFAs (save demonstratives and possessives ending in –e, which are common everywhere, although yet still more common in Croatian) are rarely used by ekavski (Serbian) speakers (Alexander, 2006: 52). Therefore, there was little expectation of finding contradictory evidence during the tabulations of the ekavski surveys.\(^{48}\) Conversely, the crux of the investigation of ALFAs rests on the purported claim that Croats do use long(er) forms more often, broadly speaking. Therefore, keeping the Labovian Observer's Paradox in mind, a few techniques were utlitized in order to elicit more "faithful" responses from the ijecavski speakers.

Understandably, if speakers were aware that their usage of ALFAs was under evaluation, according to the Observer’s Paradox they would be more self-conscious about their usage, and consequently respond more closely in line with prescriptive grammars (which list the long(er) forms parenthetically, and thus imply that they are less standard), i.e., they would reject the long forms. In order to avoid this potential pitfall, ijecavski respondents were given ekavski surveys. Switching the ijecavski surveys out for ekavski ones served another important purpose as well; it successfully

---

\(^{48}\) The results of my ekavski surveys show that this is, in fact, a generalization that can be maintained, although any purported tendency can always benefit from an empirical checkup (corpus-based research), so to speak. However, in this case, it would likely be an unfruitful enterprise.
shifted the “position of the interviewer as a learner, in a position of lower authority than the person he is talking to” (Labov 1984:40).

While ijkavian speakers were often more willing to explain the most salient differences between ijkavian and ekavian (at times quite compulsively so), they paid less careful attention to the production of ALFAs in the responses, and consequently, have supplied a wide-variety of ALFA usage. Below are the results of the questionnaire.

4.4. Results of the ALFA Questionnaire

(i) Breakdown of respondents (71 total):

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serbs living in Serbia</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>Serbs living in Croatia</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Serbs from Bosnia</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>Croats from Croatia</td>
<td>13</td>
<td>17</td>
</tr>
<tr>
<td>Montenegrins(^{49})</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 4.1 Respondent Breakdown according to Ethnicity and Gender

The youngest speaker was 18, and the oldest was 79. Montenegrin is not well represented, unfortunately, due to a lack of access to its native speakers (by my not

\(^{49}\) For a discussion of Montenegrin dialects and potential literary language, see Greenberg (2004: 88-108). Some linguists have proposed using the term BCMS, in order to be fair to Montenegrin.
having traveled through Montenegro). Nonetheless, a complete picture of BCS ALFA would include all the BCS dialects.

4.4.1 Breakdown of ALFA Usage for non-Croatian Respondents

Given that the G ALFAs –ог, –ога and the DL –ом, -ome, -omy are considered to be in free variation, the highest possible variation could only be a 33-33-33 ratio for the DL and a 50-50 ratio for the G. Although –ом and –ог are widely known to be the most common forms, respectively, let us assume they have ideal equal distribution, and a speaker that has both –е and –u only has one of the two in competition with –ом. Therefore, from a statistical perspective, we can say that the highest possible variation could be 50–50 for both G and DL ALFAs. This ratio is used as the basis for the expected totals in the chart below, where the observed column reports the number of instances of ALFA usage for each category of respondents:

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Observed</th>
<th>Expected</th>
<th>Total Possible ALFAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serbs Living in Serbia</td>
<td>13</td>
<td>136.5</td>
<td>273 (21 speakers)</td>
</tr>
<tr>
<td>Serbs Living in Croatia</td>
<td>3</td>
<td>58.5</td>
<td>117 (9 speakers)</td>
</tr>
<tr>
<td>Bosnians</td>
<td>5</td>
<td>65</td>
<td>130 (10 speakers)</td>
</tr>
</tbody>
</table>

Table 4.2 ALFA Results for non-Croatian Respondents

On the whole, non-Croatian respondents use the G and DL ALFAs in –огa and –ome, -omy only 21 times out of a possible 520 times, a mere 5% rate. Therefore, it

50 There were speakers who exhibited variation with –ом, -ome, and –omy, although this was rare.
can be concluded that the use of –oga and –ome, -omu is extremely rare in Serbian ekavian, irregardless of the place of residence of the respondent.

Interestingly, respondents who identified as Serbs but live in Croatia rarely, if ever, use ALFAs (p <.005), where it is thought to be more common. One 28-year-old Serb who had lived his whole life in Virovitica, Croatia, commented that sentence #4) *Hajde da razgovaramo o novome filmu* and sentence #5): *Videla je našega novog profesora*51 could be heard in Croatia, but that he does not use them. This was the typical response for the majority of my Serbian respondents from Belgrade as well.

One 24-year-old Serbian woman from Virovitica claimed that many of the sentences with ALFAs could be heard “u selu” (‘in the village’). Another 29-year-old Serbian woman from Virovitica claimed *tomu* (in sentence #9) was equally “Bosnian” and “Croatian”. However, this same respondent claimed that the majority of sentences with ALFAs in –oga and –ome or –omu could be heard in Croatia.

On the Bosnian side, all respondents identified as ethnic Serbs, although they have lived most of their lives in Sarajevo. Similar to the Serbs from Croatia and Serbia, the majority of Bosnian Serbs commented that ALFAs in –oga and –ome or -omu are a Croatian phenomenon. Like the 24-year-old Serb from Virovitica, one 30-year-old Bosnian woman claimed that the ALFAs were “village-speak” (again *u selu*).

For Serbs living in Serbia, respondents overwhelmingly commented that the ALFAs in –oga and –ome or –omu were Croatian, used by *neobrazovani ljudi* ‘uneducated’ people, *u selu* ‘in the village’, or by the *starija populacija* ‘older population’. Another common response was that some of the sentences would be used by *stranci* ‘foreigners’, and this returns us to the question at the beginning of Chapter

---

51 See Appendix B for translations.
3 about how much of an effect the presentation of ALFA in grammars has had on Croatian pedagogy. In this case, if foreigners are incorrectly learning ALFA usage due to the confusing and conflicting tendencies presented in grammars, one could see why a Serbian speaker might think ALFAs are a distinct mistake non-native speakers of BCS make. Interestingly, one 25-year-old male from Belgrade considered many of the sentences to be characteristic of the Serbs of Macedonian.52

4.4.2 Results for Croats living in Croatia

<table>
<thead>
<tr>
<th></th>
<th>Observed</th>
<th>Expected</th>
<th>Total Possible ALFAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Croats Living in Croatia</td>
<td>119</td>
<td>195</td>
<td>390</td>
</tr>
</tbody>
</table>

Table 4.3 Results for Croats Living in Croatia

When one compares the above table to the totals for Serbs and Bosnians, it is clear that the Croats use long forms much more often, at approximately a 33%-66% ratio, although they still predominately prefer ALFAs with only the G and DL bases –og and –om. Therefore, while not in “perfect” free variation, ALFAs in –oga, –ome, and –omu can decisively be said to be characteristic of Croatian speech (p <.005).

Curiously, Croatian respondents who did not prefer the G and DL ALFAs –oga, –ome, and –omu repeatedly located these allomorphs somewhere u Srbiji ‘in Serbia’. One 36-year-old woman from Split commented that many of the examples are

52 This is interesting given that Macedonian lacks nominal inflectional morphology. One might expect Serbs speaking a language that lacks nominal inflection to not expand on their native language’s morphology (ALFA), but to be influenced to some degree by the lack of nominal inflection in this language.
characteristic of Serbian *ekavica* ‘ekavian’. Her use of the term *ekavica* is a good example of how speakers can group another dialect according to a certain feature (here the development of CS *jat*’), but incorrectly associate this group with a feature that is not characteristic of their dialect. This is reminiscent of the discussion, in Chapter 2, of American English speakers’ perceptions about the dialectal distribution of *pop* vs. *soda* and *firefly* vs. *lightning bug* (Preston 2004).

### 4.4.3 ALFA Usage for Croatian Men vs. Women

<table>
<thead>
<tr>
<th></th>
<th>Observed</th>
<th>Expected</th>
<th>Total Possible ALFAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Croatian Males</td>
<td>68</td>
<td>84.5</td>
<td>169 (13 speakers)</td>
</tr>
<tr>
<td>Croatian Females</td>
<td>51</td>
<td>110.5</td>
<td>221 (17 speakers)</td>
</tr>
</tbody>
</table>

Table 4.4 Results for Male vs. Female ALFA Usage

The data here shows significant difference between ALFA usage for male and females. A particularly oft-quoted principle in sociolinguistics, that women are more “careful” in their speech than men, might offer insight if considered in the above distribution along sex-based lines. Labov (2006:266), on this principle, stated:

Perhaps the broadest and most widely instantiated sociolinguistic generalization concerns the careful behavior of women with stable sociolinguistic variables. It can be stated as *Principle 2, the linguistic conformity of women: For stable sociolinguistic variables, women show a

---

53 (a) Chi-square for the data from Croatian women= 33.39 (Yates),/ (P < .005), signifying that the difference between the data for Croatian men and women is meaningful.
lower rate of stigmatized variants and a higher rate of prestige variants than men.

This principle has been instantiated numerous times since the inception of sociolinguistics. For example, probably the first work to reveal varying patterns of usage among men and women was Fischer (1958), which showed that women used English the prestige pronunciation of /ing/ [ŋ] more than male speakers, who more commonly used [n], e.g. /running/ vs. /runnin'/. Lin (1988) revealed that use of retroflex consonants in Taiwanese Mandarin by women (less often by men) was a movement away from less colloquial to more formal speech. According to Labov (2004: 269), probably the most illustrative example of the principle that women speak more carefully is Eiskovits (1981). Taking 7 grammatical variables that consisted of standard and non-standard variable, Eiskovits’ data showed that girls showed a categorical preference for the standard forms for 6 of 7 variables, whereas boys consistently chose the substandard forms. Many other studies have been conducted, producing similar data and conclusions, subsequently giving validity to the notion that women prefer standard, non-stigmatized variants in their speech.

Returning to Table 4.4 above, some of my speakers commented that they do, in fact, view ALFAs as stigmatized variants. Some of my Croatian speakers who did not show preference for the ALFAs attributed them to the village, or to Serbia, demonstrating that, for them, G –oga and DL –ome, -omu are somewhat stigmatized (or perceived as being characteristic of the “other”, and not necessarily in a pejorative sense) variants. This would inevitably mean that ALFA base forms, -og and –om are prestige variants. The irony of viewing the ALFA base forms as prestige forms is that
the base forms have been said to be “puristic forms introduced after 1991 into
Croatian” (Greenberg 2004: 148). Perhaps the 51 ALFA forms chosen by Croatian
women suggests that they are straddling the fence, so to speak, on whether ALFAs are
stigmatized or prestige, and therefore in the case of ambiguity, they lean towards being
more careful.

Despite this interesting male vs. female dichotomy, at times respondents
showed tremendous intraspeaker variation. For example, one speaker, an 18-year-old
woman from Virovitica, rated Sentence #12 *Ona živi na drugomu spratu, a ja na
prvom* a ‘3’ out of 4. Yet, Sentence # 15 *Pričali smo o njenomu bratu, a ne o
njegovome* she rated a ‘4’ out of 4. While in Sentence #12 it appears she was unhappy
with the ALFA in –*omu* preceding a noun, in Sentence #15 it was deemed well-
formed, and she even accepted the addition of –*e* to *njegovom*.

4.5. Concluding Remarks

Earlier investigations (Hraste & Živković 1949, Browne 2006) have provided
contextually-based tendencies of usage for ALFAs which can no longer be said to be
true for all of Contemporary BCS. Instead, these ALFAs can now only be attributed to
Serbo-Croatian circa 1950, or Contemporary Croatian (as a puristic form), although
the wide variation in usage speaker-to-speaker suggests that all the above purported
tendencies may no longer be tenable at all in Croatian. This explains the conflicting

Furthermore, statistical analysis has proven that Croatian men use ALFAs
essentially at random, whereas Croatian women are much more careful in avoiding
them, keeping in line with Labov’s *Principle 2*, that women typically filter out stigmatized forms from their grammar. The idea that these forms are stigmatized is justified by speakers’ perceptions (from Bosnia, Croatia, and Serbia) of ALFAs as “village-speak”, characteristic of the “older generation”, or from a different BCS-speaking nation. Interestingly, the 9 Serbian speakers from Virovitica, Croatia accepted only 5% of ALFAs, and some commented that they believe the ALFAs to be Croatian. This would give additional support to the idea that the forms are evaluated in a three-way fashion: to some they signify uneducated speakers, to others an old-fashioned way to talk, and still to others “the other” (specifically Croatian vs. Serbian). The fact that the 9 Serbian speakers living in Croatia have attached this sense of the “other” to ALFAs, and thus chosen not to incorporate them into their speech is a sign that for non-Croats the most significant factor in ALFA usage is ethnicity.
CHAPTER 5

AN EMPIRICAL STUDY OF ALFA IN BOSNIAN AND CROATIAN

As discussed earlier in Chapter 3, it has been stated by Ronelle Alexander (2006: 52) and Wayles Browne (p.c. 2008) that in all of BCS the DL masculine/neuter adjectival ending in –omu (coexisting with –ome, and –om) is more common in Croatian dialects. This appears to have been verified by my sociolinguistic questionnaire. Moreover, although Serbian clearly prefers the short forms in –og (G) and –om (DL), and thus can be excluded from this corpus analysis, the Bosnian ALFA situation is not so straightforward. According to Naylor (1978: 4670), the Bosnian language, based on the Sarajevo dialect, is “western [i.e. Croatian] in phonology but closer to the eastern variant [i.e. Serbian] in morphology and vocabulary.”

Furthermore, Greenberg (2004: 148) suggests that the recent Croatian-Bosnian political unity has led to the spread of Croatian forms (as opposed to Serbian ones) in Bosnia:

Both Halilović and Jahić, the two primary advocates of a distinctive Bosnian language, have in their own writing styles shown a preference for “Croatian” forms including some of the puristic forms introduced after 1991 into Croatian… [e.g.] the marked “Croatian” forms: stoljeće ‘century, uopće ‘at all’, bit će ‘it will be’, prošlome ‘(dat.) past’, bosanskoga ‘(gen.) Bosnian’

54 The long form adjectives are extremely rare in Serbian, making the question of ALFA usage one of interest only for Bosnian and Croatian (ijekavian).

55 Wayles Browne in reaction to Pennington 2008.
(cf. the Eastern variant forms vek, uopšte, biće, prošlom, bosanskog). This choice may be a pragmatic one, since the Bosniacs\footnote{The term Bosniac is used to denote a member of the Muslim population of Bosnia.} and Croats have been joined together in a Federation in post-Dayton Bosnia-Hercegovina. This conscious decision to favor Croatian forms may also be a political hint of the anti-Serbian feelings among the Bosniac population.

The Croatian forms prošlome and bosanskoga would have been unexpected in the Bosnian of Naylor’s description, inasmuch as, in Naylor’s assessment, Bosnian patterns more with Serbian morphologically. However, Naylor was describing Bosnian of pre-collapse Yugoslavia, and therefore the recent preference for ALFAs gives more weight to Greenberg’s suggestion that there are pragmatic (in this case politically motivated) reasons for speakers choosing ALFAs in Bosnia. Nonetheless, these are the preferences of only two Bosnian speakers, Halilović and Jahić, and although they are the two most prominent advocates of the Bosnian literary language, the corpus analysis of Bosnian in this chapter can demonstrate whether the preferences of the few have become the accepted norm for the many.

Fortunately, corpus analysis of Bosnian and Croatian is quite practical nowadays, mainly due to the widespread availability of major language sources (such as the Wikipedia text dumps used in this study) and online national corpora. Moreover, thanks to the insights of Browne, Alexander, Naylor, and Greenberg, the linguist has a good starting point for empirical inquiry. Such inquiries can be useful, inasmuch as they can either bolster or contest long-held assumptions about language.

To begin, I present the materials and methodology used in carrying out this study. Then I present data from my corpus-based research on Bosnian and Croatian ALFAs. Finally, I discuss the implications of my analysis.
5.1 Materials Used in This Study

Currently, there are three online corpora for Bosnian and Croatian—the Oslo Corpus of Bosnian Texts (OCBT), the Croatian Language Repository (CLR), and the Croatian National Corpus (CNC), respectively. The OCBT has a total word count of approximately 1.5 million, whereas the much larger Croatian corpora contains upwards of 100 million (and growing). For comparison with the data mined from these online corpora, I also use entire Wikipedia database entries for both Bosnian and Croatian, which can be uploaded and queried in the R Project for Statistical Computing, a program which is currently widely used as a tool to conduct corpus analysis independent of online corpora. R also has modelling tools built into its interface, allowing the linguist to go straight from data to representations of these data in charts, graphs, etc., making it an invaluable tool for any empirically based study.

Before discussing the results of my corpus research using Wikipedia, in §5.2 I discuss the methodology for my research; in § 5.3 I discuss the query results of the corpora; and in §5.4 I give my conclusions based on the query results.

5.2 Methodology/Query criteria

Below is a list of all queries I conducted in the online corpora:

(a) Type counts for each ALFA (-ome, -omu, -om)\(^{57}\)

(b) Distribution of ALFAs for possessive/demonstrative adjectives vs. descriptive adjectives, e.g. tvom(e,u) vs. o dobrom(e,u)

(c) Sequences of ALFAs, e.g. o mom(e,u) dobrom(e,u) ‘about my good (one)’

\(^{57}\) All operations were also performed for the G –og/–oga.
Unfortunately, it is impossible to search the OCBT and CLR for individual morphemes. However, it is understood by BCS linguists that –om is by and large the most commonly occurring form in Bosnian, Croatian, and Serbia (Alexander 2006: 53). Therefore, the results for –om are not important to this study, except as a control group to –ome and –omu.58 With a much a smaller concordance report (a list of all tokens in sentential context) for –ome and –omu, I was able to scan the results to look for the distribution of –ome and –omu in the environments listed above.

5.3 Results of Corpus-based analysis

5.3.1 Total tokens for –om, -ome, -omu in the Corpus of Bosnian Texts

The initial query for –om returned 29,579 tokens, a number too large to create a concordance report for. Unfortunately, there is no function allowing for searches at the morphemic level. However, it is possible to search the corpora by commanding the search to include two or more characters (denoted with “.*”) that precede a word-final –om(e,u). This strategy eliminates the most common monomorphemic words like dom ‘house’ or tom ‘volume’.59 I included these forms in Table 4.3.1 below to demonstrate to the reader what sort of problems one can encounter when corpus results do not

---

58 For convenience, I cite the DL morpheme as –ome and –omu, although I am well aware others might prefer to cite these morphemes as –#, -e, -u built on a root morpheme base of –om.

59 While this strategy is not ideal, it is the best way to get ALFA data out of imperfect corpora.
reflect searches on a morpheme-by-morpheme basis. Moreover, if the actual examples are looked at carefully, one’s results can be skewed considerably. Luckily, hits for –ome are entirely reliable, inasmuch as the ending –ome is unique to the DL adjectival paradigm, unlike –om and –omu, which have nominal case counterparts. In the case of –omu, it occurs so rarely that the individual nominal cases can be eliminated by hand.

<table>
<thead>
<tr>
<th>ALFA</th>
<th>-om</th>
<th>-ome</th>
<th>-omu</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total hits: 32,856</td>
<td>29,579 (90%)</td>
<td>3,078 (9.4%)</td>
<td>199 (.6%)</td>
</tr>
</tbody>
</table>

Table 5.1 DL ALFAs in the Oslo Corpus of Bosnian Texts

Table 5.1 demonstrates that –omu is, in fact, quite rare in the OCBT. However, -ome is used 9.4% of the time in DL adjectival contexts. As expected, -om is used in 90% of DL adjectival contexts. Below is a list of the most commonly occurring adjectival forms in –ome and –omu:
Table 5.2 Frequency Distribution of Most Common Adjectives with DL ALFA

<table>
<thead>
<tr>
<th>tome ‘that’ 1296</th>
<th>tomu ‘that’ 47</th>
</tr>
</thead>
<tbody>
<tr>
<td>kome ‘who’ 454</td>
<td>aerodromu ‘airport’ 29 X</td>
</tr>
<tr>
<td>svome ‘one’s own’ 188</td>
<td>komu ‘who’ 6</td>
</tr>
<tr>
<td>onome ‘that one (over there)’ 156</td>
<td>cijelomu ‘whole’ 3</td>
</tr>
<tr>
<td>ovome ‘this one (here)’ 126</td>
<td>aromu 2 ‘aroma’ X</td>
</tr>
<tr>
<td>nikome ‘nobody’ 90</td>
<td>hromu ‘handicap’ 1 X</td>
</tr>
<tr>
<td>mome ‘my’ 89</td>
<td>momu ‘my’ 1</td>
</tr>
<tr>
<td>drugome ‘different’ 71</td>
<td></td>
</tr>
</tbody>
</table>

Notice that the most frequently occurring adjectives in –ome in Bosnian are the most frequently occurring adjectives in general—the possessive and demonstrative adjectives. This suggests that, in Bosnian, the likelihood of encountering an ALFA is higher dependent on the adjectival class. Perhaps Bosnian speakers reanalyzed the -e of –ome as a morphological marker of adjectival class, inasmuch as forms in –ome have been imported from Croatian (according to Greenberg 2004: 148), and consequently were somewhat alien to native speakers of Bosnian who first encountered them. In the case of –omu, to the extent that OCBT accurately reflects spoken Bosnian, it can be said that it occurs extremely rarely, except for 478 occurrences of tomu. If speakers are beginning to experiment with –omu, it is then again not surprising that –omu could possibly be gaining ground first in the closed

---

60 The following other adjectives with –ome occur more frequently than the most frequent adjective in –omu (besides tomu): nekome ‘nobody’ 71, svakome ‘each’ 57, kome ‘who’ 41, jednome ‘one’ 34, samome ‘the very’ 30.

61 ‘X’ indicates nominal inflection.
class of demonstrative adjectives, the most frequently occurring set of adjectives available with which speakers can experiment.

5.3.2 The distribution of G –og/–oga in Bosnian

If, as Greenberg has stated, that forms like *bosanskoga* ‘Bosnian’ are being promoted by advocates of a Bosnian literary language who have close ties to Croatia, we should expect a distribution of –og/–oga similar to that of –om/–ome, roughly a 9-to-1 frequency ratio. Below are the results of a query for –og/–oga:

<table>
<thead>
<tr>
<th>ALFA</th>
<th>-og</th>
<th>-oga</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total hits: 22,075</td>
<td>16,867 (76%)</td>
<td>5,208 (24%)</td>
</tr>
</tbody>
</table>

Table 5.3 G ALFA in the Oslo Corpus of Bosnian Texts

It appears –oga is a bit better entrenched in the standard language than DL –ome. However, the shorter ALFAs can be said to pattern together in terms of frequency. The longer ALFAs are universally recognized as being used in the various ways discussed in Chapter 3 (i.e. in the absence of a following noun, or for the first of a sequence of two adjectives). Unfortunately, these various environments are nearly impossible to elicit without the ability to strip words down to the morphemic level or to search by tags in the corpora (e.g. noun vs. adjective).
5.3.3 ALFA Distribution in the Croatian Online Corpora

Once again –om generated too large (1,000,000+) a hit to retrieve a concordance report. Below are the results for all three ALFAs in the CLR:

<table>
<thead>
<tr>
<th>ALFA</th>
<th>-om</th>
<th>-ome</th>
<th>-omu</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total hits: 1,149,314^62</td>
<td>1,000,000 (87%)</td>
<td>124,837 (11%)</td>
<td>24,477 (2%)</td>
</tr>
</tbody>
</table>

Table 5.4 ALFA Distribution in the CLR

When compared to the Bosnian data in Table 5.3.1, the prevalence of forms in –omu is striking. Clearly –ome occurs much more frequently than –omu in the CLR. However, in order to be absolutely certain about this frequency distribution we can do a similar query of the Croatian National Corpus. Below are the results from the CRC:

<table>
<thead>
<tr>
<th>ALFA</th>
<th>-om</th>
<th>-ome</th>
<th>-omu</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total hits 1,660,366^63</td>
<td>1,500,000 (90.3%)</td>
<td>130,257 (7.8%)</td>
<td>30,109 (1.8%)</td>
</tr>
</tbody>
</table>

Table 5.5 ALFA distribution in the CRC

It is evident then, judging from the results of the –ome and –omu queries in the CRC, that –ome occurs approximately seven times more often than –omu. In comparison with the Bosnian data, -omu occurs twice as often in Croatian than in Bosnian. This

^62 The percentages for –ome and –omu are the highest they could be if the total tokens for –om is exactly 1 million.

^63 The highest possible percentages out of 1.5 million.
small difference could possibly be enough to explain the saliency of –omu as a
“Croatian” feature”, inasmuch as its occurring less frequently can lead to speakers to
attach a social index to the morpheme when they store it in their memory (cf. the
discussion of social indexing in §2.2).

One possible explanation for the ALFA frequency distribution takes into
account the composition of the CLR and CRC, which views the type of text as having
an effect on the query results for the total number of hits for –omu. The following is
from the CLR website on the composition of their corpora (underlined are the text
types):64

The Croatian Language Corpus consists of fundamental Croatian literature
(e.g. novels, short stories, drama, poetry), non-fiction, scientific publications
from various domains and University textbooks, school books, translated
literature from outstanding Croatian translators, online journals and
newspapers, books from the pre-standardization period of Croatian language
that are adapted to nowadays standard Croatian.

Judging from the textual composition of the CLR, one could speculate that –omu is
perhaps “higher-style” Croatian due to the majority of texts being of the academic
variety. For the sake of comparison, here is a figure adapted from the website for the
other online corpus used in this study, the CNC, on its composition:

64 http://riznica.ihjj.hr/dokumentacija/index.en.html
<table>
<thead>
<tr>
<th>Text type</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Informative texts</strong></td>
<td>74</td>
</tr>
<tr>
<td><em>Newspapers</em></td>
<td>37</td>
</tr>
<tr>
<td>daily</td>
<td>22</td>
</tr>
<tr>
<td>weekly</td>
<td>9</td>
</tr>
<tr>
<td>bi-weekly</td>
<td>6</td>
</tr>
<tr>
<td><em>Magazines, journals</em></td>
<td>16</td>
</tr>
<tr>
<td>weekly</td>
<td>9</td>
</tr>
<tr>
<td>monthly</td>
<td>4</td>
</tr>
<tr>
<td>bi-, tri-monthly</td>
<td>3</td>
</tr>
<tr>
<td><em>Books, brochures, correspondence...</em></td>
<td>21</td>
</tr>
<tr>
<td>publicistics</td>
<td>4</td>
</tr>
<tr>
<td>popular texts</td>
<td>3.5</td>
</tr>
<tr>
<td>correspondence, ephemera</td>
<td>0.5</td>
</tr>
<tr>
<td>arts and sciences</td>
<td>13</td>
</tr>
<tr>
<td><strong>Imaginative texts (fiction): prose</strong></td>
<td>23</td>
</tr>
<tr>
<td><em>novels</em></td>
<td>13</td>
</tr>
<tr>
<td><em>stories</em></td>
<td>5</td>
</tr>
<tr>
<td><em>essays</em></td>
<td>4</td>
</tr>
<tr>
<td><em>diaries, (auto)biographies...</em></td>
<td>1</td>
</tr>
<tr>
<td><strong>Mixed texts</strong></td>
<td>3</td>
</tr>
</tbody>
</table>

Table 5.6 Breakdown of text type in the CNC

When comparing the two corpora, it is immediately apparent that the CLR is predominateley oriented to academic texts, which by nature reflect Croatian norms. Rather unsettling to the corpus linguist is the category “books from the pre-standardization period of Croatian language that are adapted to nowadays standard Croatian.” This statement throws up a red flag, so to speak, to the corpus researcher,
inasmuch as it appears to represent a potential for skewing the corpus with deliberately “engineered” Croatian. Interestingly, the CNC has only selected thirteen percent of its texts from the Arts and Sciences, and in general, the majority of its texts are from periodicals and other popular literature. Interestingly, however, the language politics behind the makeup of the CLR had little effect on the frequency totals of –ome and -omu in its corpus, inasmuch as the CRC shows a strikingly similar frequency distribution to the CLR for –ome and –omu. As Alexander and Browne have claimed that –omu is more Croatian, the prevalence of –omu in both the CLR and CRC verifies their assertions. Moreover, it reflects the growing status of –omu as a distinct Croatianism not only in academic, or “high style” literature, but also in the general public literature.

5.3.4 Distribution of G –og/–oga in the OCBT, the CLR, and the CRC

Below are three tables with the token counts for –og and –oga from the OCBT, the CLR, and the CRC:

<table>
<thead>
<tr>
<th>ALFA</th>
<th>-og</th>
<th>-oga</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total hits: 22,155</td>
<td>16,867 (76%)</td>
<td>5,288 (24%)</td>
</tr>
</tbody>
</table>

Table 5.7 Distribution of ALFAs –og/–oga in the OCBT
<table>
<thead>
<tr>
<th>ALFA</th>
<th>-og</th>
<th>-oga</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total hits: 1,382,162</td>
<td>1,000,000 (72.4%)</td>
<td>382,162 (27.6%)</td>
</tr>
</tbody>
</table>

Table 5.8 Distribution of ALFAs –og/–oga in the CLR

<table>
<thead>
<tr>
<th>ALFA</th>
<th>-og</th>
<th>-oga</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total hits: 1,536,863</td>
<td>1,113,374 (72.4%)</td>
<td>423,489 (27.6%)</td>
</tr>
</tbody>
</table>

Table 5.9 Distribution of ALFAs –og/–oga in the CNC

There is little statistical difference (p < .005) in the frequency distribution of ALFAs –og and –oga in the three corpora. As was demonstrated for Bosnian in the OCBT, the G –oga appears to occur more frequently than its DL counterparts, although this may be due to the fact that the DL ALFA series is tripartite, whereas the G ALFA series is bipartite. Perhaps less competition of forms in the Bosnian G creates a more stable situation where one ending –og is distinctly standard, and the other –oga is used in a “stylistic” sense.

5.3.5 Distribution of DL ALFAs in the Bosnian and Croatian Wikipedia corpora

Given that the DL ALFAs present a more complicated situation than the G ALFAs, in that they have three members, not two, I ran a check of the results I received from querying the CLR and CNC against the same query of the Bosnian and Croatian Wikipedia corpora. I ran the Bosnian and Croatian Wikipedia databases

---

\(^65\) Again, the CLR returned too many hits for –og, therefore the percentage total for –oga is its highest possible out of 1 million.
through $R$, using the exact strategy in delimiting the CRL and CNC to return only hits that had two or more characters preceding the DL morphemes $–om/–ome /–omu$.

Below are the results of the queries.

Figure 5.1 71,449 ALFs: $–om$: 66159 (93%) ; $–ome$: 4810 (7%) ; $–omu$: 480 (.07%)
The Wikipedia queries second the conclusion based on the data from the CRL and CNC that –omu is not well attested in Bosnian. However, -ome is shown to occur at the same frequency in both Bosnian and Croatian.
5.3.6 Distribution of Croatian ALFAs according to Adjectival Class

Given the views of Alexander (2006) et al., that ALFAs in –omu are used more frequently in Croatian, the support my sociolinguistic fieldwork gives to this claim, and the results of the above corpus analysis, the problem of ALFA usage can now be narrowed to discovering the environments in which ALFAs occur most frequently. This section examines the distribution of ALFAs for the three adjectival categories: possessives, demonstratives, and descriptives. Below I present a series of graphs that show frequency distributions for all the members of the closed class of possessives and demonstratives, and the six most frequently occurring descriptive adjectives. For this section I use exclusively the CNC, inasmuch as the results have been shown to be consistent with the CLR and Wikipedia. Furthermore, the CNC is largest of the three corpora. Below are a series of graphs that demonstrate the distribution of DL and G ALFAs according to adjectival class:
In Figure 5.3, there appears to be consistency of preference for the G and DL ALFAs
–og and –om. Only moj ‘my’ and svoj ‘one’s own’ exhibit relatively high returns for D
–ome (18% for moj, 20% for svoj).

---

Moj, tvoj, and svoj also have so-called longer (uncontracted forms), which are also in competition
with the ALFAs. Mojem occurs 454 times, mome 451 times. This extra competition may influence a
stabilizing reaction by native speakers in choosing one predominant form.
Figure 5.4. Distribution of Croatian G ALFAs for Possessive Adjectives

Figure 5.4 demonstrates that –og and –oga are essentially in free variation for *moj* and *svoj*. The unexpectedly high acceptance of ALFAs for *moj* and *svoj* in the G, but not in DL may suggest (as suggested earlier in §5.3.4 for Bosnian) that the DL situation is more complicated due to its comprising a set of three morphemes, instead of two. *Svoj* ‘one’s own’ occurs at such a high rate because it is used when referring to the possession of the subject of a sentence (e.g. *Marko vidi svoju devoju* ‘Marko sees his girlfriend’), thus multiplying the potential occurrences of *svoj* by six. The 2nd person possessives *tvoj* ‘your’ and *vaš* ‘your (pl.)’ occur at such low rates because the CNC is mostly composed of non-dialogical texts written in the 1st and 3rd persons.

---

67 *Mojega* only occurs 310 times, and *svojega* 1982 times, both much less than the top two competing forms, suggesting speakers prefer the contracted forms for the genitive
Immediately striking here is the sudden spike in forms in –ome, specifically in the case of tome and onome. Thus far the only speculation one could offer here is that DL tom is monosyllabic, and perhaps there is a preference for adjectival forms to be polysyllabic (at least bisyllabic). We will return to this idea later in the chapter.

---

68 I have included kakav, ovakav, takav, and onakav in the the class of demonstratives arbitrarily. They are formed from the same “spatially-related” roots as the demonstrative pronouns, but they pattern like descriptive adjectives morphologically. Moreover, these forms prefer the DL and G invariable short forms, e.g. G ovakva, DL ovakvu.
Figure 5.6 Distribution of Croatian G ALFAs for Demonstrative Adjectives

Figure 5.6 demonstrates the overwhelming preference for the long form ALFAs in – og *a* for *ovo, to*, and the much more common onoga for *ono*. At this stage, one can only speculate, but if we consider the fact that G ALFAs are much better entrenched in Croatian, then perhaps a preference for original toga, based on its monosyllabic base tog, spread via analogy to the other demonstratives.
The one surprising find in Figure 5.7 is the poor returns for kog (only 80), as opposed to the prevailing form koga. Moreover, kom and kome occur, 4641 and 4751 times respectively, suggesting equal distribution for these two forms.

69 This chart includes the “leftover” non-descriptive adjectives. I am assuming that the rest of the ordinals pattern like prvi ‘first’. The interrogative što has only long forms in the G and DL, čega and čemu, respectively. One hit turned up for kojeme, which breaks the morphotactic constraint in BCS on two consecutive e’s.
Figure 5.8 is included in this series of ALFA graphs for the sake of comparison with the other adjectival categories. The data do not contradict the finds of the CRL, the CNC, or the Wikipedia queries, which show the predominance of G –*og* and DL –*om*, with –*oga* approximately occurring twice as frequent as –*ome* in relation to its base.

Therefore, the following conclusions can be drawn about the relation of Croatian ALFAs with adjectival categories.

1) In the category of possessive adjectives *moj* ‘my’ and *svoj* ‘one’s own’ appear to demonstrate free variation for G ALFAs, but show a much higher preference for the base forms *mom* and *svom* in the DL.

---

70 The choice of these five adjectives is based on a Croatian word frequency list found at: [http://www.lexiteria.com/word_frequency/croatian_word_frequency_list.html](http://www.lexiteria.com/word_frequency/croatian_word_frequency_list.html) In Appendix C, I present my own frequency list from the CNC, which verifies the online list.
2) The demonstratives *to* ‘that one’ and *ono* ‘that one (over there)’ show preferences for DL –*ome*, whereas in the G all three demonstratives, including *ovo* ‘this one’ exhibit the predominance of forms in –*oga*.

3) The other adjectival categories, ordinals, relatives, interrogatives, and descriptive, do not deviate from the expected distribution of G and DL ALFAs found in the queries of the CLR and the CNC. Only *ko* shows an categorical preference for G –*koga*, whereas for the DL, –*om* and –*ome* are to be in free variation.

4) The only adjectives that contradict the findings of the CLR, the CNC, and the Wikipedia corpora have monosyllabic bases, e.g. *to* ‘this one’, *ko* ‘who’, *moj* ‘my’, *svoj* ‘one’s own’. Each of these forms either prefer almost exclusively G forms in –*oga* (*to, ko*) or approximate equal distribution (*moj, svoj*). *To* is the only adjective to exhibit an overwhelming preference for long form ALFAs.

5.3.7 Consecutive ALFAs in Croatian

This section is essentially an investigation of Hraste & Živković (1949), Browne (2006), who share the view that ALFA usage is restricted to the first adjective in a sequence of two, and Raguž (2010), who states the opposite, i.e. that the second adjective in a sequence of two can add a long ALFA. Below is a graph showing the results for a query for double ALFAs.
Figure 5.9 ALFAs for Two Consecutive DL Adjectives in Croatian

The CNC returned 36,010 hits for a query of two consecutive ALFA combinations. The combination –om –om occurred 30,155 (83%), signifying that Croatian speakers overwhelmingly prefer to avoid ALFAs altogether in consecutive DL adjectives. However, as the data demonstrates, speakers have experimented with different combinations, to the exclusion of –om –omu, which does not occur at all in the CNC.
Interestingly, the examples with –*omu*–*omu* are primarily from academically oriented texts. Below are a few examples:

(1) Najbolji su u onim dijelovima u kojima autor daje oduška svomu estetskome senzibilitetu

The best [times] are in those sections in which the author gives an outlet for his aesthetic sensibilities [my translation—JP]

(2) Od prvih pisanih spomenika koji se nesporno mogu pripisati srpskomu jezičnom korpusu…

From the first written manuscripts, which can uncontroversially be attributed to the Serbian language corpus… [my translation—JP]

(3) Ranoštokavski se spomenici ne mogu jednostavno i jednoznačno podijeliti po suvremenomu nacionalnomu “ključu”.

Early štokavian manuscripts cannot simply and unambiguously be divided along the modern national “key”. [my translation—JP]

(4) Bosanski Muslimani ili Bošnjaci svoj nacionalni postanak vežu za Osmansku vlast i odanost islamskomu kulturnocivilizacijskomu krugu

Bosnian Muslims, or Bosnjaks, connect their national origins with the Ottoman Empire and adherence to the Islamic cultural-civilization sphere. [my translation—JP]

(5) Ironija može biti konačna u tome da neće Hrvati i Srbi, koji su “dali” ime dvoimenomu jezičnomu hibridu[…]

The irony may be demonstrative in that Croats and Serbs don’t want [it], who “gave” a name to the double.Named linguistic hybrid. [my translation—JP]
It is noteworthy that many of the above examples have either national or linguistic themes. Moreover, the mere presence of consecutive ALFAs in \(--omu\) suggests that, although the most comfortable solution for speakers in the written language is to choose \(--om \--om\), speakers can choose to distinguish themselves by opting for ALFAs (and as I have stated earlier (cf. Introduction) and discuss further in Chapter 7, I believe this is a ethnically and/or politically motivated pragmatic decision).

Figure 5.10 ALFAs for Three Consecutive DL Adjectives in Croatian

Sequences of \(--om \--om \--om\) occurred 1,184 times, and as expected \(--ome \--om \--om\) was the second most frequently occurring combination. However, it is remarkable,
although they occur quite rarely, that in sequences of up to three adjectives almost all possible combinations of ALFAs are attested. The only combination not occurring at all in the CNC is –\textit{om} –\textit{om} –\textit{om}.

Next, I present the results of a similar corpus analysis of two and three consecutive G ALFAs:

![Figure 5.11 G ALFAs for Two Consecutive Adjectives in Croatian](image)

Not surprisingly, -\textit{og} –\textit{og} is the preferred combination in the CNC. However, out of 142,715 occurrences of two consecutive G ALFAs, -\textit{og} –\textit{og} makes up only two-thirds (down 16\% from the DL base forms in –\textit{om} –\textit{om}) of all the examples (at 67\%), while the other third is made up of ALFA combinations that prefer only one ALFA in the string.
The query for sequences of three consecutive G adjectives returned 5,860 total hits. As expected the most frequent combination, at 73%, is simply three consecutive adjectives with just the ALFA base –og. However, the other 27% is made up of a variety of combinations of G ALFAs, once again demonstrating that variation for G ALFAs is a bit better entrenched than variation in the DL ALFAs.

In conclusion, Raguž (2010) presents an inaccurate picture of sequences of multiple ALFAs. In fact, the corpora support Hraste & Živković (1949) and Brown (2006) in that when variation occurs with consecutive ALFAs speakers are more likely to choose the combinations –oga –og or –ome(u) –om, rather than the opposite, as Raguž suggests.
5.3.8 ALFAs and Preceding Prepositions

Inasmuch as G and DL contexts oftentimes require a preposition (e.g. od + G ‘from’, iz + G ‘from’, na + L ‘on’, po + D ‘according to’, etc.), an analysis of the potential effects that prepositions have on ALFAs needs to be conducted, if only to rule out this variable. In this section, I examine frequency distributions for all possessive, demonstrative, and descriptives adjectives with and without a preceding preposition. The following prepositions require the gentive case:


The following prepositions require the DL case:

k(a) ‘to, towards’, na ‘on’, nasuprot ‘opposite to’, o ‘about’, po ‘according to’, prema ‘towards’, pri ‘near, with’, u ‘in, at’

Below are a series of charts that illustrates the results of my CNC queries:
The results do not differ drastically from what we would expect based on the results illustrated by Sections 5.3.3—5.3.5. The next step then is to break the analysis down to adjectival category and monosyllabic vs. polysyllabic roots (e.g. *ko* ‘who’, *to* ‘that’, etc.), given that these were the only variables that demonstrated deviation from the overall expected patterns. Four charts are presented below, one each for possessive, demonstrative, a third group of adjectives that do not fit neatly into any of the three categories, and finally descriptive adjectives.
If compared with Figure 5.14, the distributions look very similar despite a preposition preceding the G adjective. Interestingly, however, although they are not included in the above chart, the longer forms *mojega*, *tvojega*, *svojega* occurred 62, 10, and 262 times respectively, as opposed to 30, 0, and 265 for *mojeg*, *tvojeg*, and *svojeg*. It would appear then that –*oga* is more commonly heard, proportionally speaking, in uncontracted forms than in contracted ones.

Figure 5.14 G ALFAs for Possessive Adjectives with a Preceding Preposition
The distribution for to/toga ‘that’ is the first real sign of an environmentally conditioned usage of ALFAs. When a G preposition precedes to, toga occurs 17,294 times in the CNC, as opposed to only 4,473 times for tog; an 80% to 20% ratio. However, we do not see the dominance of –oga for the other demonstratives, although there is considerable competition among both forms. The idea that adjectival class means less and syllabic structure means more in explaining ALFA usage becomes more likely once we have discovered a variable that matters: ALFAs preceded by a preposition. Now, let us turn to the “other” non-descriptive adjectives.
As suspected, monosyllabic *ko* shows an overwhelming preference for *koga*, which occurs 731 times in the CNC, and not *kog*, which only occurs 13 times—a 96% to 4% ratio. In the whole CNC corpus, *kog* occurs only 80 times, but 67 of these occurrences are without a preceding preposition—an 83% to 17% ratio. One potential pitfall for an argument that syllabic structure (monosyllabic vs. polysyllabic roots) lies at the heart of ALFA distribution are *møj, svoj*, and *koj*, which show a high rate of variation, but ultimately with forms in –*og* occurring more frequently regardless of a preceding preposition.

Finally, let us turn to the descriptive adjectives, which are all polysyllabic, and therefore should show a distinct preference for forms in –*og*, not –*oga*. 

Figure 5.16 G ALFAs for other non-descriptive adjectives with a Preceding Preposition
As suspected, the distributions for descriptive adjectives demonstrate that the presence of a preceding preposition has no effect on the occurrence of G ALFA for descriptive adjectives.

5.3.9 The Somewhat Frozen Phrase *zbog toga* ‘because of, because’

The phrase *zbog toga* occurs 9773 times in the CNC, while *zbog tog* occurs only 365 times, just 4% of the total possibilities for the phrase. However, when the phrase *zbog tog(a)* is followed by *št*o ‘that’, altogether meaning ‘because’, *zbog tog* occurs only 6 times in the CNC, whereas *zbog toga što* occurs 3,430, a 99.8% total. This practically guarantees that whether the meaning is ‘because’ or ‘because of’, *zbog toga* will be selected by speakers. Moreover, since *zbog toga*, *št*o has a strict
syntactic function in linking a main clause with a subordinate clause (e.g. *Ja sam bolestan zbog toga što nisam spavao dobro već tri dana*. ‘I’m sick because I haven’t slept for three days’), speakers have frozen the form over time, probably due to the fact that, when variation in the phrase began, speakers preferred the ALFA *toga* in the absence of a following noun, when preceded by the preposition *zbog*. Given that the modern situation shows a preference for *G –oga*, it is reasonable to believe that *zbog toga* reflects the original form, as it would seem rather contradictory that the inverse could have occurred. The 6 cases of *zbog tog* presumably have arisen due to analogy to the optionality of *G –oga*, in general.

5.3.10 The Distribution of DL ALFAs *–om, -ome, -omu* when Preceded by a Preposition

Next, I present the data for the same adjectival categories for DL ALFAs. If a frequency distribution in favor of *–ome* is as high for monosyllabics when preceded by a preposition as is the case for *–oga*, then it can definitively be said that the ALFAs pattern similarly and thus can be considered the same phenomenon for different grammatical cases.
Figure 5.18 DL ALFAs for other Possessive Adjectives with a Preceding Preposition

For the possessive adjectives, despite svoj having 2,848 hits for –oga, DL ALFAs are rather infrequently used. However, it is interesting that the only real variation occurs with the monosyllabic moj and svoj, although it the variation is not as salient as for G ALFAs occurring after a preposition. Let us now look at the demonstratives.
The CNC returned 73,531 total hits for *to* ALFAs. At 60%, *tome* occurs almost twice as often as *tom*. This is approximately the same distribution that was found for DL demonstrative ALFAs without a preceding preposition. Below are the rest of the non-descriptive adjectives:
Figure 5.20 DL ALFAs for other non-Descriptive Adjectives with a Preceding Preposition

Out of 2,727 total hits for ko with a preceding DL preposition, kome occurred 1,801 times, or 66%. Now, when this distribution is compared with the distribution of kome and kom when they are not preceded by a preposition (as shown Fig. 5.7), the results are interesting. Without a preceding preposition, kom occurs 4641 times and kome occurs 4751 times, approximately an equal distribution. The fact there is a substantial dropoff in the occurrence of kom when preceded by a preposition suggests that prepositions do, in fact, have an effect on G and DL ALFA choice for ko.
Figure 5.21 DL ALFAs for Descriptive Adjectives with a Preceding Preposition

The above illustration solidifies the previous findings in this chapter that ALFAs are rarely used with descriptive adjectives.

5.3.11 Distribution of ALFAs according to Type of Literature

As was demonstrated in §5.3.3, despite the inclusion of only 13% academic and governmental policy texts in the makeup of the CNC (which is mainly composed of periodicals and popular literature), the ALFA distribution did not differ significantly from the CLR, a corpus that is made up of solely academic texts.\(^{71}\)

However, one type of literature that is not included in either corpus is the transcription of foreign television shows and movies. Given that foreign television shows and

---

\(^{71}\) See Appendix B for a short example of a Croatian academic text. Note the repeated use of ALFAs by the author.
movies are so popular in Croatia, this type of text could be potentially insightful if analyzed as a corpus. The next analysis only looks at the overall numbers for ALFAs in a 160,000 word corpus of subtitle transcripts for foreign television shows and movies in Croatia. Below is a list of the shows and movies that make up the corpus:

<table>
<thead>
<tr>
<th>Television Show (TS) /Movie (M)</th>
<th>Total Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Parenthood- Series (S) 1, Episode 5 (S) (TS)</td>
<td>5,056</td>
</tr>
<tr>
<td>2 Sixgun (M)</td>
<td>9,438</td>
</tr>
<tr>
<td>3 Free Willy- Escape from Pirate’s Cove (M)</td>
<td>12,530</td>
</tr>
<tr>
<td>4 The Deep- S 1, E 3 (TS)</td>
<td>7,249</td>
</tr>
<tr>
<td>5 Ice Castles (M)</td>
<td>10,624</td>
</tr>
<tr>
<td>6 Ancient Aliens (M)</td>
<td>14,240</td>
</tr>
<tr>
<td>7 Wrong Turn at Tahoe (M)</td>
<td>9,806</td>
</tr>
<tr>
<td>8 The Pillars of the Earth- S 1, E 1 (TS)</td>
<td>5,705</td>
</tr>
<tr>
<td>9 The Penguins of Madagascar (M)</td>
<td>10,397</td>
</tr>
<tr>
<td>10 The Book of Eli (M)</td>
<td>8,679</td>
</tr>
<tr>
<td>11 Confucius (M)</td>
<td>12,941</td>
</tr>
<tr>
<td>12 Spartacus: Blood and Sand, S 1, E 1 (TS)</td>
<td>7,165</td>
</tr>
<tr>
<td>13 Alice in Wonderland (M)</td>
<td>8,563</td>
</tr>
<tr>
<td>14 Shutter Island (M)</td>
<td>11,108</td>
</tr>
<tr>
<td>15 The Gates- S 1, E 10</td>
<td>5,279</td>
</tr>
<tr>
<td>16 Barbie in a Mermaid Tale (M)</td>
<td>8,721</td>
</tr>
</tbody>
</table>

Table 5.10 List of TV Shows and Movies Chosen for the Corpus
The above shows and movies were chosen at random from the website:
http://www.podnapisi.net/, which has thousands of subtitles for shows and movies available as .txt files. Below are the results of a query for G and DL ALFAs:

![Chart showing distribution of ALFAs in Croatian Subtitled Foreign Films and Shows](chart.png)

**Figure 5.22 Distribution of ALFAs in Croatian Subtitled Foreign Films and Shows**

G ALFAs occur 2,371 times, with –oga occurring only 399 times, for a percentage of 17%, slightly lower than what the corpus queries of the CLR, CNC, and Wikipedia returned. However, the majority of forms in –oga are toga (80), koga (60), and moga (69), the monosyllabic stems. Interestingly, there were only two hits for svoga, as opposed to 26 for svog.

---

72 The responsible translators usually include their contact information at the end of each transcript. I suppose this would be an interesting way of getting native speaker commentary on their own ALFA usage.
As for the DL ALFAs, –ome occurred at 9.5%, with –omu occurring a negligible 10 times. If –omu is acquiring a distinctly Croatian “feel”, it is surprising that it would not be more prevalent in Croatian translations of foreign films, inasmuch as national identity is a central element in distinguishing a Croatian translation from a Serbian one, for example. The subtitling of foreign film and television would appear to an optimal conduit through which to channel a puristic or prestige form. However, its weak attestation in the above corpus suggests that –omu has not “caught on”, so to speak, in the visual media.

5.4 Conclusions of Corpus Analysis of G and DL ALFA

From these results, we can draw some conclusions about ALFAs:

1) Adjectival class is a variable that matters for Croatian speakers when choosing from the ALFAs, inasmuch as they occur rarely for the most frequent descriptive adjectives.

2) Two patterns of variation have been elicited from the CNC: a) For the closed class of possessives, the majority of demonstratives, and other non-descriptives (e.g., relatives, ordinals), the ALFA base forms are generally preferred, but svoj ‘one’s own’ and moj ‘my’ had high returns for G –oga and DL –ome. koji ‘which’ had high returns for G kojega and DL kojemu. This can perhaps be seen as a stabler form of variation (between two members, not three). Expressed differently, one could link the relatively frequent occurence of DL forms in -omu for soft-stem adjectives with the variation for G ALFAs according to how many allomorphs are competing in each category. Given that soft-stem adjectives disallow the addition of DL ALFA –e with their ALFA base
forms due to the morphotactic constraint on consecutive /e/’s (e.g. *naš ‘our’, *moj ‘my’) they only have one other form in competition with their DL base form –*om, –*omu. Having two allomorphs to choose from as opposed to three is intrinsically a more stable form of variation.

3) A generalization that emerges from the data is that monosyllabic adjectival roots prefer G –*oga and DL –*ome at much higher rates than polysyllabic ones. This particular characterization has not been stated before in the literature (see Chapter 3, where there is no mention), yet it is clear from the data for *ko ‘who’ and to ‘that’, which demonstrate an overwhelming preference for G –*oga and DL –*ome. *moj ‘my’, while *svoj ‘one’s own’, and *koj ‘which’ all exhibit some variation.

4) A preceding preposition effects the distribution of DL ALFAs for at least one BCS adjective *ko ‘who’. This is demonstrated by a significant drop in total occurrences of *kom when preceded by a preposition when compared to an approximately equal distribution of *kom ~ *kome when not preceded by a preposition. The phrase *zbog *tog(a) ‘because of’, shows an overwhelming preference for for –*oga, at 96%.

5) DL –*omu is not common in the subtitling of foreign film in Croatian, suggesting that it is not a stable variant in the language style typical of visual media.

The above conclusions play a major role in the discussion of the diachronic aspect of BCS ALFA in the next chapter, which draws on issues raised in Chapter 3 in my historiography of the presentation of BCS ALFA in Croatian and Serbian.
grammars, the findings of the sociolinguistic questionnaire in Chapter 4, and the above corpus linguistic work on the environments in which ALFAs are preferred.
CHAPTER 6

THE DIACHRONIC ASPECT OF BCS ALFA

6.1 Proposed Explanations for the Rise of BCS ALFA

As has been discussed in detail above, BCS ALFA consists of two distinct desinences for the G case, -og and –oga, and three for the DL case, -om, -ome, and –omu. The forms with three segments can be viewed as having an optional vowel added to the base forms ending in a consonant. A major historical development that may have led to the rise of DL ALFAs is the general DL syncretisms of the ancestor languages of modern-day BCS. By the end of the 15th century, the majority of Štokavian dialects had undergone a syncretism that occurred first in the nominal paradigm (Belić 1969: 10), and then spread to the adjectival paradigm.

As far as can be ascertained from the major historical grammars of Serbo-Croatian (Belić 1962/9, Marković 1967, et al.), the rise of -ome can be considered an innovation, which, according to Belić (1969: 122), is first attested for the doublet forms tom ~ tome in Eastern Štokavian, dating to the 15th century. However, these same historical grammars give unsatisfactory explanations of the origin of the innovation, and most have little insight to offer on the topic, in general. Moreover, some of the explanations given for ALFA origins in these grammars leave the reader
with more questions than answers about the nature of these variable endings, specifically about \(-e\).

Particularly striking is Marković’s attempted phonetic explanation for the rise of \(-e\) (1967: 122):

Zamjenice moj, tvoj, svoj, i koji imaju u gen. i dat. dva oblika: duže i sažete-mojega i mojemu dalo je putem progresivne asimilacije i sažimanja moga i mome...

The pronouns ‘my’, ‘your’, ‘one’s own’, which had two forms in the genitive and dative: long and short— mojega and mojemu—gave way to a progressive assimilation and contraction to the forms moga and mome. [my translation- JP]

Let us entertain Marković’s assimilation rule for *momu*, by formalizing it in linear terms:

\[
V [-\text{back}, [-\text{high}]] \rightarrow V [+\text{back} [-\text{high}]] / V [+\text{back}, [-\text{high}]] C ___
\]

(i.e. mojemu > *møjumu (later, > *møjome > mome)).

Marković assumes that the (backing) assimilation occurred before the contraction, therefore G *mojega > *mojoga > mòga (the long falling intonation is a result of contraction). While this assimilation rule seems to explain *moga, it does not explain *mome. Unfortunately, he neglects to explain the step *momu > *mome, which, in his theory would need another assimilation (a lowering), but also a fronting (a dissimilation?), to get u > e. Moreover, it is not necessary to propose an assimilation at all, as the forms mojemu and mojega still exist in modern BCS, although they are more common in Croatian (Alexander 2006: 52), and the forms *mojoga and *majome/mojumu are unattested (although this could be an accident of history).

These and other criticisms of Marković (1967) are summarized below:
1) No such historical assimilation rules (i.e. $u > e$ via lowering assimilation (and then some) have ever been a part of South Slavic historical phonology. Therefore, is hard to see any phonetic precedents for this assimilatory “vowel harmony”.

2) Marković neglects to mention that *momu* does occur, and it is listed in most grammars of Serbo-Croatian up until 1977 (the year of his grammar). How could an assimilatory rule that should affect all qualifying sequences in the language spare DL *momu*? In Marković’s approach, this could only be a later restoration of the long, uncontracted form *mojemu* (which also did not assimilate in certain contexts?), inasmuch as *momu* would have been eliminated through *mojome* $<*$mojomu*.

In essence, then, he gives an account of how particular forms might have arisen, but actually says nothing about the existence of the variation. Needless to say, Marković’s approach leaves much to be desired in terms of the regularity of sound change and a complete understanding of the history of BCS phonology.

The second, I believe, unsuccessful, but still unquestioned analysis is that of Belić (1969: 122), which posits the agglutination of a particle –*e* to the regular inflexional ending –*om*. Unfortunately, Belić did not specify what sort of “particle” he had in mind. He views the instrumental forms *nnome* ‘by me’ and *tobome* ‘by you’ as originally sequences of *nnom* and *tvom* plus a particle –*e*, which originated in the form *re* (ibid: 164), a demonstrative-like particle (we can only presume, as Belić does not explicitly state which particle he has in mind) that developed as a rhotacized
counterpart from že. This is plausible, given that ALFAs were first used for demonstratives and possessive pronominal morphology. However, many problems arise when considering the ALFA phenomenon as a whole.

How can the reduction re > e be accounted for? For example, umreti ‘to die’ and mraz ‘frost’ both have preserved clusters of /m/ and /r/. In this view, the reduction re > e is DL ALFA specific, a purported development the Neogrammarians would surely be unhappy with. Moreover, Belić overuses this particle, so to speak, to explain the origins of other problematic -e’s, whose history he is uncertain of, but which have perfectly regular historical explanations.

For example, Belić invokes this particle to explain the three-way distinction četir ~ četire ~ četiri ‘four’ in BCS (Belić 1969: 177). If re was fused to četire, how could Belić link it to the function it served for DLI singular adjectives? In fact, četire is most likely inherited from PIE *kʷetwores, which would have had an original e in the final syllable. Moreover, Vasmer cites the –e as being masculine in Proto-Slavic and -i as the feminine and neuter form: “*četyre m., *četyri ž., cp. p”. Forms of four with –i in the final syllable occur all over Indo-European as well, e.g. Lith. keturi, Skt. catvāri (ntr.), Ir cethir, etc. Less likely, but still more plausible than Belić’s suggestion, would involve Slavic remodeling its final syllable on the basis of Greek τέσσερες ‘four’ (possibly carried over from the copying of Greek Biblical texts). Even though this last statement is a stretch, it is no worse than the claim that the emphasizing particle re is responsible for the structure of the final syllable of BCS četire.
Although Belić’s idea is interesting, it cannot be taken seriously, inasmuch as he appears to apply re as a “cure-all”, so to speak, to two cases of unexpected –e, the –e being the only thing the two cases have in common. Nonetheless, one could still entertain the possibility that the DL –e originated from the fusion of a different particle to the DL ALFA base –om.

The most common type of particle that agglutinates to other lexemes in Slavic is a deictic particle. For example, Russian etot ‘this’ is formed on the Slavic demonstrative to ‘this’ (< CS *tob), which itself gave rise to the reduplicated form tobo ‘this one (masc.)’, eventually giving *e-tob-tob. The e arose as an extra deictic element to emphasize the already deictic to. However, such a view of origins of the DL ALFA –e, rasies several questions:

(1) What would motivate the agglutination of a deictic particle with an oblique case ending, or with any other inflectional ending for that matter? A survey of Slavic historical grammars shows no such development in any Slavic language.

(2) BCS does have deictic elements, such as ov- in ovde, the on- in onde, etc. However, these deictics have arisen in lexemes where an added “demonstrative” sense is semantically not surprising (for spatial relations), cf. English ‘this here’, ‘that there’, ‘that over yonder’, etc.

(3) In BCS there are no parallel cases of –e being used in an irregular, possibly deictic sense.

(4) Russian e- is prefixal, while in BCS the DL –e is suffixal.

---

73 Deixis as defined by Levinson (2000: 54): “The single most obvious way in which the relationship between language and context is reflected in the structures of languages themselves, is through the phenomenon of deixis...[Deixis] has as prototypical or focal exemplars the use of demonstratives, first and second pronouns, tense, specific time and place adverbs like now and here, and a variety of other grammatical features tied directly to the circumstances of utterance.”
Where Marković fails to explain the origin of BCS ALFA by creating an ad hoc phonological rule, which does not hold up under scrutiny, Belić simply states that –e is a particle without supporting his argument or explaining why it would have fused exclusively to oblique adjectival endings or how why re was reduced to e in only DLI ALFA contexts. Unfortunately, while both these explanations of ALFA origins ultimately come up short, they are attempts nonetheless. It is surprising that no grammar has even attempted to take these earlier accounts to task and to provide its own historical account of BCS ALFA.

Therefore I do exactly this in the next section. I first present an account of the rise of –e in the DL of the adjectival declension that stems from paradigmatic analogy (which has been proffered by Wayles Browne).\(^74\) Then I present my own account, which not only factors in paradigmatic analogy, but also considers factors such as prosody and syntax, which I believe ultimately point to the most illuminating answer yet advanced on the origins of ALFA in BCS.

6.2 The Rise of Non-Etymological ALFA –e via Analogical Leveling to the Nominal Paradigm

In the first view for the rise of ALFA -e, while the dative and locative cases were still in the process of syncretizing in the 14\(^{th}\) century, there would have been two masculine/neuter (hereafter: masc./ntr.) endings to choose from, L -e and D –u. As syncretism of the DL ran to completion, speakers chose –u as the solitary marker for both DL forms of masc./ntr. nouns. Therefore, -e would have made its way into the DL adjectival declension via analogy to the nominal paradigm (although it is

\(^74\) From personal communication with Wayles Browne, reacting to Pennington (2008).
unattested before the 15th century). At first glance, this seems to be a clean, straightforward explanation for the rise of unexpected –e. In such an account there are essentially only two steps:

1) –e still exists as a marker of locativity in the 14-15th cc..

2) When DL syncretism is completed in favor of -u, the new DL adjectival declension essentially “rescues” –e by appropriating it from the nominal paradigm.

However, while this simple two-step explanation seems optimal, after a cursory glance, I believe it leaves too many questions unanswered. Although, as Labov has famously stated, historical linguistics is the “art of making the best use of bad data” (Labov 1994: 11), and therefore tends to prefer as few assumptions as possible in formulations of historical reconstructions, in the case of BCS ALFA, a view that only looks to nominal inflexion does not consider the role of other elements in BCS inflexion, such as syntax and prosody. Moreover, I believe this account leaves out many important details about the diachrony of modern-day BCS that should be at least factored in or decisively ruled out before a well-informed hypothesis can be formulated. Therefore, I believe the above account, which views the nominal L –e as the source for ALFA –e, is untenable.

First, when DL syncretism ran to completion in the 15th century, Eastern Štokavian (the ancestor to modern-day Serbian) underwent the complete loss of –e as a marker of nominal dativity/locativity, with ultimately the D -u diachronically prevailing. Given that ALFA origins date only to the 15th century, there would have been a narrow gap for speakers to map a waning marker of locativity into the DL
adjectival paradigms. Moreover, it is even more puzzling to consider why any language would need three distinct markers of dativity/locativity for adjectives. Furthermore, keeping in mind that –e has no clear etymology for either the D or L adjectival paradigms (CS *D –omy, L *-omy), it seems highly unlikely that the –e of –ome was originally related at all to nominal or adjectival morphology. Therefore, one must look outside the nominal paradigm for related phenomena in the grammar to help elucidate the origins of BCS ALFA.

One can identify various reasons why D –u won out over L –e, e.g. 1) partially because D –u was the desinence for both the D nominal (-u) and D adjectival paradigm (-om-u), and 2) the preposition u ‘in’, as a separate, syntactic, element in locative phrases, e.g. u gradu ‘in the city’ (earlier L *u grade) may have allowed for some “sound” association (homophony?) with the D morphological –u during the height of DL syncretism.75 Such a view for the rise of nominal DL –u does not just simply consider a one-to-one morphological correspondence, i.e. nominal: adjectival, but also examines peripherally related, non-nominal elements.

Another criticism of the proposed nominal-based account is the very existence of DL adjectival –e. It is curious why –e is needed at all when –u can do the work, so to speak, for both dative and locative contexts. Following Kuryłowicz’s (1966 [1945-1949]: 169) 4th Law of Analogy,76 in languages where two forms compete to fulfill the same function, the innovative form commonly is assigned the everyday meaning at the

---

75 Before the 15th century neo-štokavian stress retraction, the preposition u would not have been homophonous in all environments, in particular when an end-stressed noun like kralj ‘king’ (Slavic accent paradigm B) would have had stress on its dative ending in –u, e.g. kraljú. Prepositions in Slavic are historically enclinomena (underlyingly unstressed). However, if a noun had so-called oxytonic (accent paradigm C, which showed mobility in the root) stress, a preposition could take default stress.

76 In all fairness, this “law” is only a tendency and is not meant to be predictive (cf. Hock 1991, Chapter 11).
expense of the conservative form, which is relegated to a secondary function (e.g. English *brothers ~ brethren*). With both –*e* and –*u* occurring as ALFA variants in the same contexts this archaic/innovative split is not apparent. They *both* were used, according to the grammars listed in Chapter 3, originally as markers of definiteness (the innovative usage of ALFAs), at times in the absence of a noun (*mudromu je i jedno oko dosta* ‘For the wise man, one eye will suffice’ (Hraste & Ţivković 1949:96)) and at times for entire adjectival phrases (*upuštite se zelenome travnjaku* ‘set off in the green meadow’ (Stevanović 1964: 262). However, with all the confusion in the literature about how ALFAs have been used, it is difficult to explain how exactly the roles of the ALFAs have shifted over time, thus making the modern-day situation all the more bewildering.

An additional criticism of a nominal-based account is that it does not look outside the nominal paradigm to other inflexional paradigms in the grammar where parallels of long (DL -*omel-omu*)/short forms (DL -*om*) already exist. While the locative would originally have been short –*om*, in terms of segmental length, the dative was always long –*omu*, going back at least to Common Slavic (every Slavic language has a dative case adjectival form in –*omu*). In making a long form from –*om* (and not from the already “long” pre-existing –*omu*) speakers may have looked to where the distinction of long/short already exists, namely in the pronominal paradigm, which does, in fact, show evidence that –*e* was just as strong a marker of locativity as it was of dativity for certain dialects of Serbian during the critical period of DL syncretism. Belić notes that in “southern dialects” of historical Serbian (he believes these to be Montenegrin), the accusative, genitive, dative, and locative forms of ‘me’,
‘you’, and ‘oneself’ were all syncretic, e.g. *mene/me, tebe/te, sebe/se* (Belić 1969: 164). Karadžić lists these forms as well in his 1824 grammar, although he does not specify in which regions they were heard in his time.

### 6.3 An Integrated Hypothesis for the Rise of Unexpected –e and its Spread to Croatian Dialects

As DL syncretism ran to completion in Serbian (and had likely already occurred in Croatian), -e began to be added to DL forms of the adjectives. Some modern grammars (Hraste & Živković (1949)/ Browne (2003/6)), have described the ALFAs as adding the sense of definiteness to an adjective, but this might only reflect just a fairly modern development, inasmuch as BCS historically marked adjectival definiteness through morphology, vowel length, and accent.

There appears to have been a suprasegmental morphonological divide between indefinite/definite adjectives in Vuk Karadžić’s (1824) grammar. In the 15th century, Eastern Štokavian underwent a stress retraction one syllable leftward. The resulting intonation on syllables that newly acquired stress was either long or short rising, depending on the underlying length of the syllable. One other development, the metatony of rising intonation to falling in closed syllables “balanced out” the prosodic inventory, so to speak, by increasing the likelihood of falling tones in closed syllables. From the Neogrammarian view of sound change, the retraction must have occurred in all applicable phonetic environments (i.e., where there was a stress that could have shifted leftwards), and the same logic applies to the metatony of rising to falling in closed syllables. Therefore, it appears an adjective like žút ‘yellow-indef.’, for
example, which shows a long fall in the nom./masc./sg. as a result of the earlier Common Slavic neo-acute retraction (stress retraction off of a short vowel or a *jer*), e.g. nom.sg.masc. žút < *žút < *žutó, píješ < *píjejó ‘to drink (2nd sg.)’, ultimately shows this later development of metatony of rising tones in closed syllables. However, Vuk lists all the other “indefinite” forms of žút with a long rise on the initial syllable, indicating that indefinite forms would have been end-stress originally, e.g. fem. žúta < *žutá. Conversely, the definite adjective had word-initial stress, e.g. žúti, which is an open syllable and exhibits a long falling intonation in all forms. Belić (1969: 154) observes a similar (seemingly related) intial-stress (definite) vs. end-stress distinction:

…u većem delu štokav[skog] govoru svugde u neodređenom vidu, sem u jednosložnim oblicima, akcenat razvio na kraju, a u određenom na početku. Nema sumnje da je za to bio od značaja i tip: bēo [white], bēla, bēlo: bēlī, bēlā, bēlō.

…for the most part, in the štokavian dialect, everywhere in the indefinite form, except for in monosyllabic forms, word-final accent developed, but for definite forms, word-initial. There is no doubt the following forms are relevant in this regard: bēo [white], bēla, bēlo: bēlī, bēlā, bēlō. [my translation—JP.]

Belić’s examples support my claim that indefinite and definite forms were in a supersegmental complementary distribution, involving word-initially stressed forms, i.e. definite, and word-finally stressed forms, i.e., indefinites. I disagree, however, with Belić’s claim that indefinites were end-stressed in all but monosyllabic forms. The indefinite form bēo presumably would have originally been end-stressed, e.g. belí.

---

77 This is sometimes referred to as “Stang’s Law”, after Christian Stang, who first formalized it in his Slavonic Accentuation (1957).

78 This, in principle, could match Russian žéltý, -ogo, -omu, etc., the first syllable of which has fixed-initial stress. The initial-stress in Russian is typically explained synchronically by the general lack of stress on long (attributive) adjectival endings (unless the adjective is idiosyncratically end-stressed, like bol’sój, for example). However, the feminine short form (predicative) žel’tá is perhaps evidence of stress mobility in the root, which would have been subsequently reanalyzed as attributive vs. predicative instead of definite vs. indefinite.
before the loss of jers in weak positions. Then the subsequent neo-acute stress retraction from the *jer*, and the change of *l > o* in Štokavian (pre-15th century) would explain the short falling intonation of *bêo*. An account of definiteness in BCS that factors in the neo-acute stress retraction can explain the seemingly odd falling intonation in *bêo* and *žut*, and, moreover, demonstrates that definite vs. indefinite appear to demonstrate an accentual complementary distribution.⁷⁹

With an established phonologically (vowel length), prosodically (accent), and morphologically (for the nominative masc. sg.) marked definiteness in 15th century Eastern Štokavian, it seems surprising that the ALFAs would have only been acquired by the DL and G, and not the instrumental case for adjectives, for example. Furthermore, the fact that the ALFA endings persisted despite the potential that they might split according to Kuryłowicz’s 4th Law of Analogy, makes BCS ALFA all the more interesting. Moreover, one must ask just how many ways of marking definiteness does a language need, anyway.

Therefore, I believe it is important to not simply focus on morphological leveling from a strict form-to-form sense, inasmuch as looking there for an explanation leads to more questions than answers. However, an account which factors in other elements of morphology (morphosyntax and prosody) as well as principles of contact linguistics, can shed light on the more curious aspects of BCS ALFA. From the prosodic side, if the pronominal paradigm of 14-15th cc. Serbian is factored in, the diachronic morphological aspect regarding where –*e* originated and the diachronic

---

⁷⁹ Subsequent analogical leveling has caused a lot of confusion with accentual paradigms in BCS, but this is outside the scope of this dissertation. In the future, a large-scale diachronic study of definite vs. indefinite adjectives in BCS from the accentual perspective could shed light on whether my purported complementary distribution existed.
syntactic aspect, as to the syntactic environment in which the ALFA was added, becomes much more transparent. Once I establish an explanation for the rise of unexpected –e in Serbian in § 7.3.1, I then explain how it spread to Western Central Slavic dialects (modern-day Croatian) in § 6.3.2., drawing on principles of contact linguistics.

6.3.1 New Hypothesis for the Rise of Unexpected ALFA –e in Eastern Štokavian

Below I list the four major points, which restate exactly where the diachronic side of ALFA stands up to this point in the discussion: 1) At the time when supposedly the nominal –e would have been mapped into the adjectival paradigm (14-15th cc.), DL syncretism was moving towards completion in favor of –u. For this to be the source, one would have to assume the spread of a waning marker of nominal dativity/locativity in –e to spread to another paradigm of declension (adjectival), which already has two available endings, -om and –u, to fulfill a function which has yet to be explained from a diachronic perspective in historical grammars. More plausible is a view that considers two possible pressures that were exerting influence over the morphological development of the DL case in Eastern Štokavian, the stability of –u as a marker of dativity, which has been preserved in all contemporary Slavic languages in the D nominal declension, and, 2): u ‘in’ as a preposition requiring a following noun in the DL, e.g. u gradu ‘in the city’, which may have helped speakers associate u’s generally with locativity, thus propelling it to the forefront of DL syncretism at the expense of –e.
The third point from section 6.2 considers the pragmatic issues with having three functionally equivalent allomorphs: Why would speakers of Eastern Štokavian need three inflectional morphemes for the DL if morphemes are to be considered “units whose function is, precisely, to enable speakers to say different things” (Garcia 1985)? Another element that has not been considered in discussions on the origin of the BCS ALFA phenomenon is the suprasegmental facts.

The fourth point is that the simple nominal: adjectival morphological account does not look outside the nominal inflectional paradigm for possible parallels of long/short forms in the language (e.g. pronominal paradigm, auxiliaries). If we consider that definiteness in adjectives was originally decided by suprasegmental factors, then suprasegmental factors in BCS ALFA should be ruled out first when attempting to explain subsequent developments. It is particularly puzzling and yet unexplained what the function of ALFA was initially, if not an expression of definiteness. However, it appears there is strong evidence which supports a hypothesis which views suprasegmental factors as integral to the development of BCS ALFA, aside from the foggy diachronic semantics. Particularly strong cross-paradigmatic evidence can be discovered by looking to the pronominal paradigm.

First, many of the Slavic languages exhibit stressless elements, enclitica, which “move” to second position in a phrase (the so-called “Wackernagel” position). Historical South Slavic inherited short, enclitic forms for its long pronouns (meni/mi, tebi/ti, etc.) from Proto-Indo-European, which are canonically found in 2nd position. Similarly, short forms for the copular biti ‘to be’ and the future auxiliary from hteti ‘to want’ were formed: e.g. jesam/sam and hoću/ću respectively. If pronouns and verbs
could gain an additional grammatical function (for pronouns the added element is “focus”) through their integration into second-position prosodics, speakers could have integrated adjectives into a new type of “mini” 2nd position prosodics, namely when occurring immediately before an omitted noun, in order to gain the new grammatical function of definiteness. This is a complicated argument, and therefore I lay out the details step-by-step below.

Let us first consider two points: 1) one of the main environments for ALFA usage is when a DL adjective stands alone without a noun to modify (Hraste & Živković 1949: 96), and 2) DL and G phrases are oftentimes preceded by prepositions, which are historically enclinomena in Slavic. If points 1) and 2) reflect the original usage of ALFAs in 15th century Eastern Štokavian, one could imagine a scenario where speakers adapted so-called “2nd position” to other parts of sentence structure that can be considered independent constituents, namely prepositional phrases. The fact that prepositions are enclinomena is important, given that they could never be stressed unless they took default stress from following stressless elements (adjectives and nouns of accent paradigm C). If a preposition gives up its 1st position status, so to speak, to a following adjective, then the position immediately following an adjective would inherit 2nd position status in this view. In the case of ALFA prepositional phrases, this element would be the noun. Once prepositional phrases began “mirroring” broader phrasal structure, incorporating their own Wackernagel-like position, in the absence of a following noun, it appears speakers replaced this “null” 2nd position with a stressless element that could replace a noun. These stressless

---

80 This is along the lines of Baker (1985), who proposed a “mirror principle” that sees parallels in syntactic and morphological organization. For a particularly illustrative example of the “mirror principle”, cf. Nevis & Joseph (1992) on the Wackernagel-like mobile reflexive affix -s/i).
elements, ALFAs, would have taken an adjective as a host, just as pronominal and auxiliary clitics “lean” on a leftward host. This is the critical step in which speakers may have looked to the pronominal paradigm to find a suitable element to fill this null 2\textsuperscript{nd} position (note that this view redefines 2\textsuperscript{nd} position in terms of syllabic structure and not syntactic/prosodic structure).

Secondly, if the unexpected nature of DL adjectival –e is taken into account, the pronominal declension again looks to be a much stronger “donor”, so to speak, of the ALFA in –e.\textsuperscript{81} In fact, Belić (1969) lists menel/me, tebe/te as dative and locative long and short forms in his historical grammar. Furthermore, Vuk Karadžić (1824) lists mene/me, tebe/te as forms for the accusative, genitive, dative and locative. Moreover, contemporary modern Montenegrin and Croatian dialects still show syncretism of the DL 1\textsuperscript{st} and 2\textsuperscript{nd} person pronouns in –e (Belić 1969). As a strong marker of dativity/locativity in 15\textsuperscript{th} century Eastern and Southern Štokavian, where Belić claims ALFAs originated, the DL pronominal morpheme -e could have been appropriated by speakers as a nominal short form, or anaphoric pronoun, thus fulfilling the semantic role of definiteness and the prosodic role of filling a null 2\textsuperscript{nd} position, in the absence of a noun.

A direct parallel can be seen in pronominal prepositional phrases, such as od njega ‘from him’ and o meni ‘about me’, but *od ga, and *o mi, where long form pronouns are obligatory. If this constraint reflects 15\textsuperscript{th} century Eastern Štokavian, it could have spurred the formation of similar prepositional adjectival phrases. Therefore, in prepositional adjectival constructions such as o tom (psu) ‘about that

\textsuperscript{81} Admittedly, while I have been concerned with previous historical accounts of ALFA which reconstruct forms unattested in BCS, my argument would be more cogent were the forms men, teb, and seb attested.
(dog)’ and od tog (stola) ‘from that (table)’, the preposition, as a proclitic forces the adjective into 1st position. Then o tom and od tog mirror *od ga and *o mi, which then add –e or –u and –a, respectively, in order to fulfill an apparent requirement on the length of prepositional phrases, that is as having at least two feet. Moreover, there are many examples in BCS where syntactic/prosodic 2nd position coincides with the 2nd syllable in a phrase, e.g. biću ‘I will be’, piću ‘I will drink’, Ja ti kažem ‘I tell you’, etc. This, in theory, explains why ALFAs are much more common for adjectives with monosyllabic roots.

Another important fact to consider from the diachronic perspective is where ALFAs are recorded first. Belić (1969) states that the long form in –e was first used with possessive and demonstrative adjectives, and we can imagine speakers then later spreading the long forms to descriptive adjectives, thus making ALFAs a general adjectival phenomenon. If we view DL –e as originating in the pronominal paradigm, then it is not surprising to see that the first place this pronominal element is used is with possessive pronouns and demonstratives. I call this phenomenon “pronominal association”. Descriptive adjectives rank last on a cline of “pronominality”, inasmuch

---

82 There are pronominal prepositional phrases where the preposition has only one syllable, e.g. o njāj ‘about her’, o njih ‘about them’, o vās ‘about you’. Notice, however, that in all these phrases, the complement forms (the pronouns) have long vowels, as opposed to their short forms, which all exhibit short vowels. It appears that these forms shows the effects of a lengthening under ictus (i.e. long form positions), suggesting that perhaps the addition of an ALFA does not reflect an original syllabic context, but moric. However, this possibility will not be explored in this dissertation, inasmuch as a “two-syllable” constraint explains the great majority of prepositional phrases in BCS. A tendency that began on moraic terms and then spread to syllabic terms can be seen in the Greek “trimoric rule”, as explained by Hamp (1989: 44): “Western Makedonski dialects show a dominant pattern, with their three-syllable rule, a generalization or maximization of the Greek positional constraint which was known in Classical Greek as the “trimoric rule”.* Just as Classical Greek generalized this accentuation for the originally unaccented Indo-European main clause finite verb, in an original bilingual contact situation less competent speakers must have applied the trimoric (= trisyllabic, since there is no prosodic vowel length in Makedonski) rule to Eastern South Slavic words, including criticiced forms, as if they had been accentless.”
as they are an open class and do not refer to “person” or “proximity”, as the
possessives and demonstratives do. Therefore, descriptive adjectives are the last
paradigm to import the ALFA –e.

Below I give a detailed step-by-step summary of my new integrated hypothesis
for the rise of ALFAs in Eastern Štokavian:

1) In modern BCS, pronominal adjectives share long form/short form distinction
with personal pronouns, although functionally different. For example, o mome
‘about my...”, or o mom both occur, whereas mene/me ‘you ACC’ is in prosodic
complementary distribution, i.e. stressed vs. unstressed.

2) Diachronically speaking, adjetival inflection originates from lexical root +
pronominal suffixes. i.e. *dobra-jego, *dobra-jemu (simplified). So there is
already a diachronic typological precedence for the incorporation of pronominal
elements as clitic-like elements, which then attach leftward to an adjetival
lexeme.

3) At first, demonstrative adjectives shared the innovation of ALFA morphology
with the possessives, suggesting that speakers considered the two to be part of a
similar grammatical class. It appears adjectives acquired ALFAs at a later period
due to being “the least pronominal” class of adjectives on a cline of pronominality,
i.e. the demonstratives and possessive adjectives are part of a “bound”
grammatical system, whereas adjectives are part of “free” lexical class, but which
operate according to the basic rules of demonstrative and possessive morphology,
i.e., they had only had the ALFA base forms in –om and –og originally.
4) At the end of the 15th century, in Eastern Štokavian, DL syncretism had completed for the singular forms of all pronouns, and most nouns (which favor –u) (Belić 1962: 8).

5) D/L/I syncretism then spreads to the plural in the same adjectival categories (Belić 1962: 10).

6) Speakers began to reanalyze instances where nouns could be omitted in adjectival phrases such as o mom (psu) ‘about my (dog)’, along the lines of how they treated prepositional pronominal phrases, which to this day require long forms after prepositions, e.g. *od ga, but od njega. Two steps would have occurred in this process: a) a new element needed to be added to lengthen monosyllabic adjectival phrases as prepositional objects in accordance with the “two-syllable” (perhaps better, “two-foot”) rule for clitics and pronominal prepositional phrases, and b) this new element was selected according to the context of the prepositional phrase, i.e. G prepositional phrases selected –a and DL selected –e or –u. Given that these forms were originally only added when a following noun was omitted, they were semantically linked to the omitted noun as anaphoric markers of definiteness.

7) With 2nd position having a systematic status for pronouns and auxiliaries, speakers looked to these systems to find the optimal clitic to replace an omitted noun, and thus fulfill the semantic function of definiteness and the prosodic function of occupying 2nd position.

8) Instead of looking towards the nominal paradigm’s locative morpheme –e, which was lost at the expense of D –u as DL syncretism came to completion in the
15th century, speakers found this –e in pronominal DL *mene/me, tebe; te*, in the sense that these forms are based on the roots, *men-*, *teb-*, and *t-*, respectively.

9) After syncretism of the D/L/I plural was completed in BCS, the new D/L sg. marker -e could have slipped into the Isg. of demonstrative, and interrogative pronouns (e.g. *time* ‘by that’, *kime* ‘by whom’, *čime* ‘by what’, etc.), but subsequently failed to spread throughout all adjectival paradigms in the Isg.

Steps 8 and 9 are further elaborated below in a Four-Part Analogy Approach:83

(i) \[a : a']
   \[b : X-b']

(ii) \[\text{dog : dog-s}
   \[\text{cat : cat-s}
   \[\ldots \ldots
   \[\text{cow } X= \text{cow-s (replacing older kine)}

As we can see in the above model, the English plural for cow, older *kine*, was replaced by the innovative form, *cows*, based on the formation of similar plural forms for *dog* and *cat*. Similarly, this type of analogy explains the spread of –e marker of dativity/locativity from long form personal pronouns to pronominal adjectives:

(iii) \[\text{mi : men-e (alternating with men-i in Dsg.)}
   \[\text{ti : teb-e (alternating with teb-i in DSG.)}
   \[\ldots \ldots

---

83 Model adapted from Hock (2005).
moj-em $\rightarrow$ móme (alternating with mómú)

(iv.) -e is leveled throughout all other adjectival categories as a variant marker of dativity/locativity in long forms

(v) Apply Kuryłowicz’s Fourth Law of Analogy

As cows replaced kine, as an everyday plural for cow, kine persisted only in high-style or Biblical English. The interesting twist for ALFAs is that they were innovative in the 15th century, but now are deemed archaic, as the majority of BCS dialects prefer the older base form ALFAs for both the G and DL. Now –e and –u have been reassigned pragmatically restricted use, especially in Croatia.

The same four-part model can be used to explain the extension of –e into the Isg. of pronouns (i.e. mnøme, tvøme, sviøme, tìme, ovøme, ćìme, kìme).

6.3.2 The Spread of ALFA –e to Croatian Dialects

An important factor that must be kept in mind when investigating historical Serbian and historical Croatian is that while Serbian is completely Štokavian, Croatian includes not only Štokavian, but also Kajkavian and Čakavian (the labels being based on the word for ‘what’ in each dialect).\textsuperscript{84} There are dialects to this day in Čakavian-speaking territories, such as Trogir, where ALFAs are entirely lacking (Geić 1998). In fact, more radically, we see GD syncretism with the G adjectival –oga winning out over D –omu in and around Trogir. Furthermore, the historic center of Croatia, Zagreb, falls under the scope of Kajkavian (which exhibits many similarities with Slovenian, such as an analytic future formation with an I-participle verb form, e.g. ja bum delal ‘I

\textsuperscript{84} For a detailed map of the BCS dialects, see Appendix F.
will do' and č, j, from PS *tj, *dj) shared a much closer affinity to Slovenian in the Middle Ages. However, through four centuries of repeated Serbian migrations to Croatia, the scope of Štokavian grew much wider, inasmuch as it was “shared” across the Austro-Hungarian/Ottoman political and geographic demarcation (to the East of which was “Serbia”, to the West of which was “Croatia”). Pavle Ivić provides the push and pull factors for Serbian migrations to Croatia (Ivić 1972: 74):

Political causes include the ravaging wars in the epoch of the Turkish invasion, pillaging expeditions of Turkish military commanders, adverse circumstances in life in the regions already dominated by the Turks, enticement of the population by the neighboring Christian states trying to organize a defense of the borderline and to colonize the regions of the Turkish side of the boundary… The general direction of migrations triggered by most of these causes was away from the Turkish territory towards that under Christian rule… Economic causes of migrations include the infertility of the soil in the highlands, overpopulation of those regions, and famine as a consequence of drought…An additional impetus came from the fact that the population of plains and valleys was often decimated by wars, plagues, emigration, etc. The general population movements brought about by such causes was from the highland to the lowland.

Clearly, the major migrations of Serbian Štokavian speakers were from East and South, where the Muslim Turks imposed their will, thus making sedentary life much more difficult, to the West and North, where their Christian brethren awaited them with literally “greener pastures”.

Due to these migrations, the linguistic consequences were “manifold”, according to Ivić (ibid: 78):

The distribution of dialectal types underwent far-reaching changes. Areas of certain dialects were substantially enlarged [including so-called ekavski, originally based in Šumadija-Vojvodina-JP]…Certain other dialect areas shrank, for instance, those of the Čakavian and Kajkavian dialects, as well as

---

85 I used quotes here to emphasize that neither country had existed independently since pre-Austrian (for the Kingdom of Croatia) and pre-Ottoman (for the Serbian Kingdom) times. I am well aware of the identity issues that the use of these two terms bring up.
of most among the western Štokavian dialects... Certain dialectal types were affected by a strong influence of another dialect whose speakers had settled in their areas or around them. E.g., some western Štokavian dialects in Slavonia and Bosnia, which originally had a number of peripheral peculiarities forming a transition, as it were, to Čakavian, became much closer to average Štokavian after the colonization of East Hercegovina je-speakers in Bosnia and Slavonia...

Not only did Štokavian speakers expand their dialect by resettlement, the close linguistic proximity of Štokavian to Čakavian and Kajkavian caused a somewhat unilateral process of contact-induced grammatical borrowings (such as da + complements for infinitives, undoubtedly a Serbian import, and the neo-Štokavian stress retraction, which did not take place in Kajkavian, Slavonian Štokavian (only partially there), and Čakavian (which still exhibits an accentual system closer to that of Common Slavic). Furthermore, 19th century political trends, particularly the “Ilyrian movement”, whose adherents sought unity for all of South Slavdom, put even more pressure on the minority dialects of Croatia to look more like, as Ivić puts it, “average Štokavian”—the dialect chosen by Vuk as the standard. As these Illyrianist sentiments grew stronger, the Vienna Literary Agreement (signed in 1850) established a standard language that demanded linguistic unity among the various dialects, preferring Vuk’s codified grammar. Furthermore, the consequent consolidation of most of the Bosnian-Croatian-Serbian speaking areas of the Balkans into the Yugoslavan state in the 20th century continued the predominance of Eastern Štokavian norms (modern-day Serbian), as the capital city was located in Belgrade.

These repeated, accumulating entrenchments of Serbs and Serbian linguistic norms in Croatia have been shown to have had a lasting effect on the native Croatian dialects. As ekavian and jekavian Serbs brought with them their dialect features
(accentual system, syntax, lexemes), it is likely they imported DL –e and the
optionality of ALFAs. One could counter this statement by claiming that DL
syncretism was a shared innovation in Serbian and Croatian dialects (which is true to
some extent), and that these migrations are irrelevant to the story of ALFA origins.
However, while Serbian and Croatian dialects likely shared a starting point for DL
syncretism, the Austrian-Ottoman divide facilitated divergent end results for dialects
on either side. This view rests on the absence of ALFAs from Croatian grammars up
until 1860 (ten years after the Vienna Literary Agreement, calling for adherence to
Štokavian norms), and the attestation of forms in –e in 15th century Serbian (Belić
1969: 122). Moreover, DL syncretism could have occurred independently in Serbian
and Croatian dialects, as it is not typologically uncommon in Indo-European
languages, in general (for Italian and Sardinian, cf. Calabrese 1995; for Ancient
Greek, cf. Coleman 1987). Therefore, in this account, Serbian is the basis for the
introduction of ALFAs in –e and –u into Croatian, yet they have the status of being
distinctly Croatian. This ironic turn of events in the history of Serbo-Croatian may
have caused the leaders in Croatian language policy to rethink their reintroduction of
the D ALFAs –e and –u in post-Yugoslav Croatian, had they known the diachronic
details.

86 It is conceivable that unattested ALFAs in Croatian before 1860 could simply be an “accident of
history”. Ancient Greek was thought for a long time to have lacked the Proto-Indo-European word for
‘sister’. However, as luck would have it, the form eor was eventually discovered, showing the regular
loss of s in initial position in Greek (Janda & Joseph 2003).
87 The feminine forms of the DL in many Slavic languages are syncretic. For a large-scale description of
case syncretism in the Indo-European languages, see Luraghi (1987).
CHAPTER 7

CONCLUSIONS

In this dissertation, I have taken a synchrony and diachrony approach, in the spirit of Joseph (1983), to examine the historical origin of Bosnian-Croatian-Serbian adjectival long-form allomorphy (ALFA). Previous accounts that viewed BCS ALFA as markers of definiteness are no longer tenable in Contemporary BCS. This has been stated by Alexander (2006:52) and verified by my sociolinguistic questionnaire, the results of which demonstrate the usage of BCS ALFA is synchronically strictly sociolinguistic in nature.

However, some sociolinguists (Milroy & Gordon 2003, Chesire 1999, Hudson 1996) hold the view that morphosyntactic elements, such as BCS ALFA, cannot inherently be sociolinguistic in nature, inasmuch as, in their view, the higher frequency of a form entails a higher potential for social indexing, given that speakers begin to attach pragmatic meaning (social information) to forms that are consistently and frequently characteristic of a certain social group. While this is undoubtedly true, I argue that critics of the sociolinguistic morphosyntactic variable do not consider that their argument is essentially based on saliency of form. Saliency can occur with either a feature that is frequent, like a Bostonian’s “r-lessness”, a phonological variable, or
something that is less frequent, like an Appalachian English speaker’s use of the
creation *needs fixed*, a morphosyntactic variable, which for a non-Appalachian
speaker can have a jarring effect on comprehension. Therefore, the views of Milroy &
Gordon (2003) et al., should be revised to consider the potential for social indexing to
exist on a frequency spectrum, such that less frequent, non-phonological (higher level)
variables have a lower critical mass for being salient, and frequently occurring
phonological (lower level) variables have a higher critical mass for being salient. In
accordance with this view, BCS ALFA has acquired a social indexing, given that they
occur infrequently in most of the dialects of BCS, and speakers of non-Croatian
dialects typically (correctly) ascribe ALFAs to Croatian speech, or they attach some
other stigmatized index.

The few grammars that have attempted to explain contextually based
tendencies of usage for ALFA have generally been conflicting, to say the least. Not
only do most grammars differ in the conventions they use to present the ALFAs, but
they also often differ in their explanations for ALFA usage. For example, Hraste &
Živković (1949) present the first account for BCS ALFA rules of usage, stating that in
a string of two consecutive adjectives, the first of the two can opt for an ALFA.
Moreover, they are the first to mention that ALFAs are commonly used in the absence
of a following noun (and functioning as markers of definiteness). However, just fifteen
years later, Stevanović (1964) challenges Hraste & Živković (1949) by claiming
ALFAs can occur on both adjectives in a string of two, or that when three in a row
occur the middle adjective cannot opt for a long form. This discussion is taken to
absurd lengths by Raguž (2010), who proffers a new hypothesis of G and DL
“epenthetic” vowels –a and DL –e, -u, which break up clusters that arise when a following word has an intial consonant that is similar to the –g of G –og (velar) or the –m of DL -om (labial), e.g. momak #kupuje ‘the boy buys’. Moreover, Raguž adds to the confusion generated by Stevanović (1964) and Hraste & Ţivković (1949) by claiming that in a string of two adjectives, the first or second can “distinguish” itself by choosing an ALFA, e.g. po starome dobrom običaju ~ po starom dobrome običaju. Ironically, Raguž has convincingly shown in his own speech that the tendencies he presents are erroneous (cf. Chapter 3).

Following up on Raguž (2010), I conducted sociolinguistic fieldwork in Croatia and Serbia in order to take to task the many non-empirically based statements of BCS ALFA that have plagued previous grammars. The results categorically demonstrate that non-Croats rarely (<5%) use ALFAs. Moreover, respondents typically considered ALFAs stigmatized, in the sense that they viewed them as elements of “uneducated”, “elderly”, or “Croatian” speech. Even Serbian respondents who have spent their entire lives in Croatia understand the Croatian nature of ALFAs, and this is the likely cause why they have not incorporated ALFAs into their speech; for Croatian Serbs, their non-use of ALFAs is a gesture towards their Serbian-oriented identities.

Once ALFAs are narrowed solely to Croatian speech, further divisions of their distribution can be discerned. ALFA usage can be viewed as another exemplary case of Labov’s “Principle 2”, which states that women are generally more careful about using stigmatized forms in their speech and tend towards a more careful style of speech, as opposed to men. Croatian women used ALFAs much more than any non-
Croatian group observed in this study. However, when compared to Croatian men, who use ALFAs frequently and haphazardly, Croatian women were consistently more careful in rejecting ALFA forms (p < .05).

Furthermore, from speaker to speaker, AFLA usage varied tremendously in my questionnaires. The contextually based environments proposed in previous grammars have been shown to no longer be tenable for Contemporary BCS. In order to reveal whether there are, in fact, any preferred environments for ALFAs, I conducted corpus analysis on a variety of questions of BCS ALFA usage.

First, the results of queries of ALFA usage in the Oslo Corpus of Bosnian Texts, the Croatian Language Repository, the Croatian National Corpus, and Wikipedia have all verified claims by Alexander (2006) and Browne (2006) that ALFA usage, especially the DL –u, is more Croatian, although G –a and D –e occur in Bosnian. Secondly, two contextual factors have a bearing on ALFA distribution in the corpora: adjectival class and the presence of preceding preposition.

In the CNC, ALFAs are rarely found for descriptive adjectives. However, interestingly, the most frequently occurring descriptive adjective with ALFAs is hrvatski ‘Croatian’. As it turns out, ALFAs occur much more frequently for monosyllabic possessives and demonstratives, the most frequently occurring with ALFAs being monosyllabic to ‘that’ and ko ‘who’, which show an overwhelming preference for ALFAs when preceded by a G or DL preposition.

As separate constituents, prepositional phrases have their own internal prosodic structure, inasmuch as they can occur either to the left or right of a main clause, e.g. BCS O mom psu sam pričao ~ Pričao sam o mom psu ‘I spoke about my
dog’. In this first variant, the entire prepositional phrase takes 1\textsuperscript{st} position, which is immediately followed by a stressless element in 2\textsuperscript{nd} position, the past tense auxiliary from *biti* ‘to be’.

Taken as constituents that have their own internal prosodic structure, prepositional phrases reveal the diachronic side of the ALFA problem. Keeping in mind that prepositions in Slavic are historically enclinomena, i.e. elements that can only bear default stress (in cases when stress is inherited from a following word that belongs to accent paradigm C, words with mobile stress), in prepositional phrases, prepositions occupy “syntactic” first position, but not “prosodic”. The preposition leans rightward on a following adjective forming one orthotonic or tactile group. No problems arise when these prepositional orthotonic groups are followed by a noun to modify, with the noun occupying 2\textsuperscript{nd} position. However, given that speakers were able to omit nouns from these phrases (if referentiality had been established), they developed a tendency to fill in these new “null” 2\textsuperscript{nd} positions. This tendency can be considered a “two-syllable constraint”, which developed originally for the monosyllabic *tog* and *tom*. In the case of the DL, instead of adopting –*e* from the nominal paradigm, where it was a waning marker of locativity, speakers brought in –*e* from the pronominal DL *mene, tebe, sebe* (which was characteristic of some 15\textsuperscript{th} century Eastern and Southern Štokavian dialects, and still characteristic of some Montenegrin dialects); the –*e* was a second-syllable element. ALFAs then conveniently functioned as anaphoric pronouns in the absence of a following noun, which explains why, since Karadžić (1824), grammars have viewed ALFAs as adding an added sense of definiteness (although definiteness was already thrice...
tautological). In a sort of “pronominal association,” speakers first applied ALFA usage to possessives and demonstratives (which are also the only adjectival categories that have monosyllabic stems after the addition of ALFA base form of G –og and DL -om), and then subsequently spread their ALFA usage to descriptive adjectives via analogy.

Finally, as Greenberg has claimed (Greenberg 2004: 148), ALFAs were reintroduced as “puristic” elements in Croatian after the dissolution of Yugoslavia. However, based on the lack of attestation of ALFAs in Croatian until the 19th century, ALFAs could only be seen as puristically Serbian and Croatian, reminiscent of the Vienna Literary Agreement of 1850 that called for linguistic unity among the Serbs and Croats. In an even more ironic twist, Serbian refugees, fleeing Ottoman occupation likely brought ALFAs with them to Croatian Štokavian, along with their neo-Štokavian stress retraction and their usage of da as a complementizer, among other features. Therefore, there is nothing distinctly “Croatian” about ALFAs at all, in the diachronic sense, although synchronically they are clearly being pushed in academic, or “high-style” literature.
APPENDIX A: A CHRONOLOGY OF ILYRIAN/CROATIAN GRAMMARS

88 Adapted from the Institute for Croatian Language website: http://www.ihjj.hr/oHrJeziku-gramatike.html


Blaž Tadijanović, *Svaschta po mallo illiti kratko sloxenye immenah, i ricsih u illyrski, i nyemacski jezik.* Magdeburg, 1761, Tropava, 1766.


Ignacij Szent-Mártony, *Einleintung zur kroatischen Sprachlehre für Deutsche.* Varaždin, 1783.

Franz Kornig, *Kroatische Sprachlehre oder Anweisung für Deutsche die kroatische Sprache in kurzer Zeit gründlich zu erlernen nebst beygefügten Gesprächen und verschiedenen Uebungen.* Zagreb, 1790, 1795.


Josip Voltić, *Grammatica illirica, u: Ricsoslovnik illiricskoga, italianskoga i nimacskoga jezika s' jednom pridpostavleniom grammatikom illi pismenstvom: sve ovo sabrano i sloxeno od Jose Voltiggi Istranina*, Beč, 1803.

Franjo Marija Appendini, *Grammatica della lingua Illirica ...* Dubrovnik, 1808, 1828, 1848, 1850.


Vjekoslav Babukić, *Osnova slovnice slavjanske narričja ilirskoga uredjena Věkoslavom Babukićem. Danica ilirska, II, 1836*

Ignac Kristijanović, *Grammarik der Kroatischen Mundart.* Zagreb, 1837.


Rudolph Fröhlich (Veselić), *Der kleine Ilirer.* Beč, 1840.


Andrija Stazić, *Grammatica della lingua illirica ad uso degli amatori nazionali e stranieri che bramano d'impararla*. Zadar, 1850.


Adolfo Veber (Tkalčević), *Skladnja ilirskoga jezika za niže gimnazije*. Vienna, 1859, 1862.


Vinko Pacel, Oblici književne hrvaštine. Karlovac, 1865.

Pero Budmani, Grammatica della lingua serbo-croata (illirica). Vienna, 1867.

Adolfo Veber (Tkalčević), Slovnica hrvatska za srednja učilišta. Zagreb, 1871, 1873, 1876.

Josip Vitanović, Slovnica hrvatskoga jezika za nižu realku. Osijek, 1872.

Ivan Danilo, Slovnica za srednja učilišta nižega reda. Zadar, 1873.


Josip Vitanović, Gramatika hrvatskoga jezika za škole i samouke. Osijek, 1880, 1882, 1888, 1891.

Mirko Divković, Nauka o izreci. Zagreb, 1880, 1892.


Mirko Divković, Oblici i sintaksa hrvatskoga jezika za srednje škole. Zagreb, 1903, 1908.


APPENDIX B: THE ACADEMIC/OFFICIAL USAGE OF ALFAs
The following text is from a review of the first Croatian grammar, Kasić’s *Osnove ilirskoga jezika* (1604). All are DL masc./ntr. adjectival forms that occur in the first three pages are in bold font (Horvat 2002: 113-115). This text is a good illustration of how academic texts really promote the use of ALFAs, although the usage is at times often quite inconsistent with regard to environment (such as the ones proposed by Raguž (2010)) or grammatical case, i.e D or L.

“Osnove ilirskoga jezika, koje su mnogi priželjkiivali, ako i manje savršene u svojim dijelovima, dovršio sam svojim snagama, doista s veseljem i predanošću, možda samo u prevelikoj žurbi. Nije se, naime, do sada pojavio nitko (koliko znam) koga bih kao pisca mogao slijediti i čija bi se pravila o ilirskome govoru pojavila u pisanom obliku piše anima koji uči ilirski jezik u predgovoru prve tiskane gramatike hrvatskoga jezika mladi isusovac s otoka Paga, Bartol Kasić. I kao što je prije gotovo četiri stotine godina, zahvaljujući ovome obrazovanom i vrijednom Pažaninu, hrvatski jezik dobio zasluženo mjesto u europskome jezikoslovlju, taka je i Institut za hrvatski jezik i jezikoslovlje napravio niz značajan korak za suvremenu hrvatsku lingvistiku objavivši pretisak i prijevod jednoga od najznačajnijih djela naše jezikoslovne, ali i kulturne baštine. Zahvaljujući kvalitetnome prijevodu s latinskoga jezika Sanje Perić Gavrančić i opširnome pogovoru Darije Gabrić-Bagarić, s kronološki početnim izdanjima Bartola Kasića te literaturuom o njemu, temelj hrvatske gramatike dostupan je svim istraživacima, ali i zaljubljenicima u
APPENDIX C: FREQUENCY LIST OF INDIVIDUAL ADJECTIVES IN CNC
WITH ALFA–OME AND –OMU
<table>
<thead>
<tr>
<th>Term</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>tome</td>
<td>63284</td>
</tr>
<tr>
<td>kome</td>
<td>4715</td>
</tr>
<tr>
<td>svome</td>
<td>4005</td>
</tr>
<tr>
<td>ovome</td>
<td>4003</td>
</tr>
<tr>
<td>onome</td>
<td>3069</td>
</tr>
<tr>
<td>nikome</td>
<td>2433</td>
</tr>
<tr>
<td>drugome</td>
<td>2083</td>
</tr>
<tr>
<td>nekome</td>
<td>1884</td>
</tr>
<tr>
<td>svakome</td>
<td>1611</td>
</tr>
<tr>
<td>prvome</td>
<td>1500</td>
</tr>
<tr>
<td>Hrvatskome</td>
<td>1322</td>
</tr>
<tr>
<td>jednome</td>
<td>1252</td>
</tr>
<tr>
<td>samome</td>
<td>993</td>
</tr>
<tr>
<td>novome</td>
<td>729</td>
</tr>
<tr>
<td>istome</td>
<td>673</td>
</tr>
<tr>
<td>nikomu</td>
<td>439</td>
</tr>
<tr>
<td>svakomu</td>
<td>437</td>
</tr>
<tr>
<td>svomu</td>
<td>434</td>
</tr>
<tr>
<td>umnogome</td>
<td>395</td>
</tr>
<tr>
<td>onomu</td>
<td>387</td>
</tr>
<tr>
<td>cijelome</td>
<td>374</td>
</tr>
<tr>
<td>drugomu</td>
<td>329</td>
</tr>
<tr>
<td>ovomu</td>
<td>320</td>
</tr>
<tr>
<td>nekomu</td>
<td>319</td>
</tr>
<tr>
<td>glavnome</td>
<td>240</td>
</tr>
<tr>
<td>jednomu</td>
<td>195</td>
</tr>
<tr>
<td>starome</td>
<td>179</td>
</tr>
<tr>
<td>takvome</td>
<td>173</td>
</tr>
<tr>
<td>radnome</td>
<td>155</td>
</tr>
<tr>
<td>velikome</td>
<td>154</td>
</tr>
<tr>
<td>Term</td>
<td>Value</td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>ikome</td>
<td>145</td>
</tr>
<tr>
<td>mnogome</td>
<td>144</td>
</tr>
<tr>
<td>javnome</td>
<td>139</td>
</tr>
<tr>
<td>malome</td>
<td>137</td>
</tr>
<tr>
<td>pravome</td>
<td>137</td>
</tr>
<tr>
<td>Svakome</td>
<td>130</td>
</tr>
<tr>
<td>samomu</td>
<td>129</td>
</tr>
<tr>
<td>kratkome</td>
<td>121</td>
</tr>
<tr>
<td>rodnome</td>
<td>120</td>
</tr>
<tr>
<td>veznome</td>
<td>116</td>
</tr>
<tr>
<td>Novome</td>
<td>107</td>
</tr>
<tr>
<td>dobrome</td>
<td>105</td>
</tr>
<tr>
<td>prvomu</td>
<td>100</td>
</tr>
</tbody>
</table>
APPENDIX D: QUESTIONNAIRES
EKAVIAN QUESTIONNAIRE

Lingvistički Upitnik
Ovo je anonimni, dobrovoljni upitnik koji od ispitanika traži da ocene rečenice ocenom od 1 do 4. U upitniku tražimo samo podatke o uzrastu, polu, mestu boravka i etničkoj pripadnosti. Ukoliko se ispitanik slaže da priloži ove podatke onda mora dati verbalnu saglasnost pre početka ispitivanja.

Godine starosti:

Pol:

Mesto i država rođenja:

Mesto i država stanovanja:

Etnička pripadnost:

---

Ocene:
(4) Upotrebio(la) bih tu rečenicu.

(3) Čuo(la) sam ovakvu rečenicu, ali ja je ne bih upotrebio(la).
Gde bi se ovo moglo čuti? Ko mislite da bi mogao upotrebiti ovu rečenicu?

Šta biste vi rekli umesto toga?

2) Ova rečenica zvuči čudno ali mogu da “zamislim” da je neko upotrebi.
Gde bi se ovo moglo čuti?

Šta biste vi umesto toga rekli? Ko mislite da bi mogao upotrebiti ovu rečenicu?

1) Ovo nikada nisam čuo(la) i ne mogu da zamislim da bih igde mogao(la) da čujem tu rečenicu.

PRIMER: Sad živimo u ovomu lepome mestu. (Upišite ocenu 1-4)________
Ako ste odabrali ocenu (2) ili (3), molimo vas da odgovorite na dodatna pitanja
1. Da li danas ideš u banku? ______
Gde bi se ovo moglo čuti? Ko mislite da bi mogao upotrebiti ovu rečenicu?

Šta biste vi rekli umesto toga?

2. Juče, bili su u novomu lepome parku. ______
Gde bi se ovo moglo čuti? Ko mislite da bi mogao upotrebiti ovu rečenicu?

Šta biste vi rekli umesto toga?

3. Da li znate ovoga slavnoga glumca? ______
Gde bi se ovo moglo čuti? Ko mislite da bi mogao upotrebiti ovu rečenicu?

Šta biste vi rekli umesto toga?

4. Hajde da razgovaramo o novome filmu. ______
Gde bi se ovo moglo čuti? Ko mislite da bi mogao upotrebiti ovu rečenicu?

Šta biste vi rekli umesto toga?

5. Videla je našega novog profesora. ______
Gde bi se ovo moglo čuti? Ko mislite da bi mogao upotrebiti ovu rečenicu?

Šta biste vi rekli umesto toga?

6. Sutra ćemo da razgovaramo o ovom. ______
Gde bi se ovo moglo čuti? Ko mislite da bi mogao upotrebiti ovu rečenicu?

Šta biste vi rekli umesto toga?

7. Nije čuo vesti o dobrome sinu nego o lošemu. ______
Gde bi se ovo moglo čuti? Ko mislite da bi mogao upotrebiti ovu rečenicu?

Šta biste vi rekli umesto toga?

8. Plesali su s mladima lepima curama. ______
Gde bi se ovo moglo čuti? Ko mislite da bi mogao upotrebiti ovu rečenicu?
Šta biste vi rekli umesto toga?
___________________________________________________________

9. Jako sam se obradavao tomu. ________
Gde bi se ovo moglo čuti? Ko mislite da bi mogao upotrebiti ovu rečenicu?

Šta biste vi rekli umesto toga?
___________________________________________________________

10. Ovog mi je stvarno dosta. ________
Gde bi se ovo moglo čuti? Ko mislite da bi mogao upotrebiti ovu rečenicu?

Šta biste vi rekli umesto toga?
___________________________________________________________

11. Nisam znao da ćete o tome da me pitate. ________
Gde bi se ovo moglo čuti? Ko mislite da bi mogao upotrebiti ovu rečenicu?

Šta biste vi rekli umesto toga?
___________________________________________________________

12. Ona živi na drugomu spratu, a ja na prvom. ________
Gde bi se ovo moglo čuti? Ko mislite da bi mogao upotrebiti ovu rečenicu?

Šta biste vi rekli umesto toga?
___________________________________________________________

13. Čim se baviš? ________
Gde bi se ovo moglo čuti? Ko mislite da bi mogao upotrebiti ovu rečenicu?

Šta biste vi rekli umesto toga?
___________________________________________________________

14. Nisu nikomu ništa rekli. ________
Gde bi se ovo moglo čuti? Ko mislite da bi mogao upotrebiti ovu rečenicu?

Šta biste vi rekli umesto toga?
___________________________________________________________

15. Pričali smo o njenomu bratu, a ne o njegovome. ________
Gde bi se ovo moglo čuti? Ko mislite da bi mogao upotrebiti ovu rečenicu?

Šta biste vi rekli umesto toga?
___________________________________________________________

156
Lingvistički upitnik
Ovo je anonimni, dobrovoljni upitnik koji od ispitanika traži da ocijene rečenice ocjenom od 1 do 4. U upitniku tražimo samo podatke o uzrastu, spolu, mjestu boravka i etničkoj pripadnosti. Ukoliko se ispitanik slaže da priloži ove podatke onda mora dati verbalnu suglasnost prije početka ispitivanja.

<table>
<thead>
<tr>
<th>Godine starosti:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Spol:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mjesto i država rođenja:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mjesto i država stanovanja:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Etnička pripadnost:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

Ocjene:
(4) Upotrijebio(la) bih tu rečenicu.

(3) Čuo(la) sam ovakvu rečenicu, ali ja ju ne bih upotrijebio(la).
Gdje bi se ovo moglo čuti? Što mislite tko bi mogao upotrijebiti ovu rečenicu?

Što biste Vi rekli umjesto toga?

(2) Ova rečenica zvuči čudno ali mogu “zamisliti” da je netko upotrijebi.
Gdje bi se ovo moglo čuti?

Što biste Vi umjesto toga rekli?

1) Ovo nikada nisam čuo(la) i ne mogu zamisliti da bih igdje mogao(la) čuti

PRIMJER: Sad živemo u ovomu lijepome mjestu. _ (Upišite ocjenu 1-4)_

Ako ste odabrali ocjenu (2) ili (3), molimo Vas da odgovorite na dodatna pitanja:
Pišite 1-4

1. Ideš li danas u banku? ________
   Gdje bi se ovo moglo čuti? Što mislite ko bi mogao upotrijebiti ovu rečenicu?

   Što biste Vi rekli umjesto toga?

2. Jučer, bili su u novomu lijepome parku. ________
   Gdje bi se ovo moglo čuti? Što mislite ko bi mogao upotrijebiti ovu rečenicu?

   Što biste Vi rekli umjesto toga?

3. Da li znate ovoga slavnoga glumca? ________
   Gdje bi se ovo moglo čuti? Što mislite ko bi mogao upotrijebiti ovu rečenicu?

   Što biste Vi rekli umjesto toga?

4. Hajdemo da razgovaramo o novome filmu. ________
   Gdje bi se ovo moglo čuti? Što mislite ko bi mogao upotrijebiti ovu rečenicu?

   Što biste Vi rekli umjesto toga?

5. Vidjela je našega novog profesora. ________
   Gdje bi se ovo moglo čuti? Što mislite ko bi mogao upotrijebiti ovu rečenicu?

   Što biste Vi rekli umjesto toga?

6. Sutra ćemo razgovarati o ovom. ________
   Gdje bi se ovo moglo čuti? Što mislite ko bi mogao upotrijebiti ovu rečenicu?

   Što biste Vi rekli umjesto toga?

7. Nije čuo vjesti o dobrome sinu nego o lošemu. ________
   Gdje bi se ovo moglo čuti? Što mislite ko bi mogao upotrijebiti ovu rečenicu?

   Što biste Vi rekli umjesto toga?

8. Plesali su s mladima lijepima curama. ________
<table>
<thead>
<tr>
<th></th>
<th>Gdje bi se ovo moglo čuti? Što mislite ko bi mogao upotrijebiti ovu rečenicu?</th>
<th>Što biste Vi rekli umjesto toga?</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.</td>
<td>Ovog mi je stvarno dosta.</td>
<td></td>
</tr>
<tr>
<td>11.</td>
<td>Nisam znao da ćete me o tome pitati.</td>
<td></td>
</tr>
<tr>
<td>12.</td>
<td>Ona živi na drugomu katu, a ja na prvom.</td>
<td></td>
</tr>
<tr>
<td>13.</td>
<td>Čim se baviš?</td>
<td></td>
</tr>
<tr>
<td>15.</td>
<td>Pričali smo o njenomu bratu, a ne o njegovome.</td>
<td></td>
</tr>
</tbody>
</table>
ENGLISH TRANSLATION OF QUESTIONNAIRE

Linguistic Questionnaire
This is an anonymous, voluntary questionnaire, which requires the respondent to judge sentences on a 1-4 scale. The only background information about respondent are age, sex, place of birth and residency, and ethnicity. If the respondent agrees to these conditions (s)he is required to first give verbal agreement before beginnin the questionnaire.

<table>
<thead>
<tr>
<th>Date of Birth:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex:</td>
</tr>
<tr>
<td>Place and Country of Birth:</td>
</tr>
<tr>
<td>Place and Country of Residency:</td>
</tr>
<tr>
<td>Ethnicity</td>
</tr>
</tbody>
</table>

-------------------------------------------------------------------------------------------------------

Values:
(4) I would use this sentence.

(3) I’ve heard this sentence, but I wouldn’t use it.
Where could you hear it? Who might use it?

What would you say instead?

2) These sentence sounds strange, but I can imagine someone saying it.
Where could you hear it?

What would you say instead?

1) I’ve never heard this before, and I can’t imagine that anyone would ever use it.
EXAMPLE: Now I live in the new, beautiful park. (Choose from 1-4)
If you chose (2) or (3), we ask that you respond to the questions provided below the sentence.

Write from 1-4

1. Are you going to the bank today? _________
Where could you hear it? Who might use it?
________________________________________________________________
What would you say instead?
________________________________________________________________

2. Yesterday, they were in the new, beautiful park. _________
Where could you hear it? Who might use it?
________________________________________________________________
What would you say instead?
________________________________________________________________

3. Do you know that famous actor? _________
Where could you hear it? Who might use it?
________________________________________________________________
What would you say instead?
________________________________________________________________

4. Let’s talk about the new film. _________
Where could you hear it? Who might use it?
________________________________________________________________
What would you say instead?
________________________________________________________________

5. She saw our new professor. _________
Where could you hear it? Who might use it?
________________________________________________________________
What would you say instead?
________________________________________________________________

6. Tomorrow we will talk about it. _________
Where could you hear it? Who might use it?
________________________________________________________________
What would you say instead?
________________________________________________________________

7. He didn’t hear news about the good son, but about the bad one. _________
Where could you hear it? Who might use it?
________________________________________________________________
What would you say instead?
________________________________________________________________
8. They danced with the young, pretty girls. ________
Where could you hear it? Who might use it?
________________________________________________________________
What would you say instead?
________________________________________________________________

9. I am so happy about it! ________
Where could you hear it? Who might use it?
________________________________________________________________
What would you say instead?
________________________________________________________________

10. I’ve really had enough of this. ________
Where could you hear it? Who might use it?
________________________________________________________________
What would you say instead?
________________________________________________________________

11. I didn’t know you were going to ask me about that.________
Where could you hear it? Who might use it?
________________________________________________________________
What would you say instead?
________________________________________________________________

12. She lives on the second floor, and I on the first. ________
Where could you hear it? Who might use it?
________________________________________________________________
What would you say instead?
________________________________________________________________

13. What are you up to? ________
Where could you hear it? Who might use it?
________________________________________________________________
What would you say instead?
________________________________________________________________

14. They didn’t tell anyone. ________
Where could you hear it? Who might use it?
________________________________________________________________
What would you say instead?
________________________________________________________________

15. We spoke about her brother, not his. ________
Where could you hear it? Who might use it?
________________________________________________________________
What would you say instead?

_______________________________________________________________
APPENDIX E: DATA TABLES FOR FIGURES FROM CHAPTER 5


<table>
<thead>
<tr>
<th></th>
<th>om</th>
<th>ome</th>
<th>omu</th>
</tr>
</thead>
<tbody>
<tr>
<td>moj</td>
<td>1,913</td>
<td>451</td>
<td>84</td>
</tr>
<tr>
<td>tvoj</td>
<td>105</td>
<td>49</td>
<td>42</td>
</tr>
<tr>
<td>njegov</td>
<td>4,316</td>
<td>108</td>
<td>7</td>
</tr>
<tr>
<td>njezin</td>
<td>1,303</td>
<td>33</td>
<td>6</td>
</tr>
<tr>
<td>naš 'our'</td>
<td>5,387</td>
<td>664</td>
<td></td>
</tr>
<tr>
<td>vaš 'your (pl.)'</td>
<td>1062</td>
<td>120</td>
<td></td>
</tr>
<tr>
<td>njihov 'their'</td>
<td>3009</td>
<td>54</td>
<td>1</td>
</tr>
<tr>
<td>svoj 'one's'</td>
<td>15,640</td>
<td>4,005</td>
<td>434</td>
</tr>
</tbody>
</table>

Figure 1  DL ALFA for Possessive Adjectives

<table>
<thead>
<tr>
<th></th>
<th>og</th>
<th>oga</th>
</tr>
</thead>
<tbody>
<tr>
<td>moj</td>
<td>1,611</td>
<td>1536</td>
</tr>
<tr>
<td>tvoj</td>
<td>72</td>
<td>214</td>
</tr>
<tr>
<td>njegov</td>
<td>3,283</td>
<td>108</td>
</tr>
<tr>
<td>njezin</td>
<td>945</td>
<td>177</td>
</tr>
<tr>
<td>naš 'our'</td>
<td>5,385</td>
<td>2,477</td>
</tr>
<tr>
<td>vaš 'your (pl.)'</td>
<td>732</td>
<td>284</td>
</tr>
<tr>
<td>njihov 'their'</td>
<td>2132</td>
<td>181</td>
</tr>
<tr>
<td>svoj 'one's'</td>
<td>13,000</td>
<td>9,905</td>
</tr>
</tbody>
</table>

Figure 2 G ALFA for Possessive Adjectives
<table>
<thead>
<tr>
<th></th>
<th>om</th>
<th>ome</th>
<th>omu</th>
</tr>
</thead>
<tbody>
<tr>
<td>ovo</td>
<td>'this</td>
<td>21,728</td>
<td>4003</td>
</tr>
<tr>
<td>to</td>
<td>'that</td>
<td>43,619</td>
<td>63284</td>
</tr>
<tr>
<td>ono</td>
<td>'that on</td>
<td>2654</td>
<td>3069</td>
</tr>
<tr>
<td>ovakav</td>
<td>'in th</td>
<td>574</td>
<td>35</td>
</tr>
<tr>
<td>takav</td>
<td>'in tha</td>
<td>3334</td>
<td>173</td>
</tr>
<tr>
<td>onakav</td>
<td>'in th</td>
<td>72</td>
<td>5</td>
</tr>
<tr>
<td>kakav</td>
<td>'in wh</td>
<td>1687</td>
<td>67</td>
</tr>
</tbody>
</table>

Figure 3  DL ALFA for Demonstrative Adjectives

<table>
<thead>
<tr>
<th></th>
<th>og</th>
<th>oga</th>
</tr>
</thead>
<tbody>
<tr>
<td>ovo</td>
<td>'this</td>
<td>26,815</td>
</tr>
<tr>
<td>to</td>
<td>'that</td>
<td>24,703</td>
</tr>
<tr>
<td>ono</td>
<td>'that</td>
<td>3144</td>
</tr>
<tr>
<td>ovakav</td>
<td>'in</td>
<td>500</td>
</tr>
<tr>
<td>takav</td>
<td>'in th</td>
<td>3388</td>
</tr>
<tr>
<td>onakav</td>
<td>'in</td>
<td>39</td>
</tr>
<tr>
<td>kakav</td>
<td>'in wh</td>
<td>947</td>
</tr>
</tbody>
</table>

Figure 4  G ALFA for Demonstrative Adjectives
Figure 5  DL and G ALFA for other non-Descriptive Adjectives

<table>
<thead>
<tr>
<th></th>
<th>og/eg</th>
<th>oga/ega</th>
<th>om/em</th>
<th>ome</th>
<th>omu</th>
</tr>
</thead>
<tbody>
<tr>
<td>koj 'which'</td>
<td>19426</td>
<td>10919</td>
<td>42699</td>
<td>1</td>
<td>15533</td>
</tr>
<tr>
<td>ko 'who'</td>
<td>80</td>
<td>5836</td>
<td>4641</td>
<td>4751</td>
<td>1494</td>
</tr>
<tr>
<td>prvi 'first'</td>
<td>9981</td>
<td>3031</td>
<td>13965</td>
<td>1500</td>
<td>100</td>
</tr>
<tr>
<td>jedan 'one'</td>
<td>18839</td>
<td>5124</td>
<td>28765</td>
<td>1252</td>
<td>195</td>
</tr>
</tbody>
</table>

Figure 6  DL and G ALFA for Descriptive Adjectives

<table>
<thead>
<tr>
<th></th>
<th>og</th>
<th>oga</th>
<th>om</th>
<th>ome</th>
<th>omu</th>
</tr>
</thead>
<tbody>
<tr>
<td>hrvatski 'Croatian'</td>
<td>19922</td>
<td>6834</td>
<td>12296</td>
<td>1322</td>
<td>132</td>
</tr>
<tr>
<td>veliki 'big'</td>
<td>5403</td>
<td>957</td>
<td>5405</td>
<td>154</td>
<td>45</td>
</tr>
<tr>
<td>glavni 'main'</td>
<td>5903</td>
<td>953</td>
<td>2728</td>
<td>240</td>
<td>9</td>
</tr>
<tr>
<td>novi 'new'</td>
<td>12294</td>
<td>3478</td>
<td>7272</td>
<td>729</td>
<td>83</td>
</tr>
<tr>
<td>mali 'small'</td>
<td>2439</td>
<td>853</td>
<td>1972</td>
<td>137</td>
<td>38</td>
</tr>
</tbody>
</table>

Figure 7  DL ALFAs for Two Consecutive Adjectives

<table>
<thead>
<tr>
<th></th>
<th>DL ALFAs and Consecutive Adjectives in Croatian</th>
</tr>
</thead>
<tbody>
<tr>
<td>ome ome</td>
<td>893</td>
</tr>
<tr>
<td>omu omu</td>
<td>237</td>
</tr>
<tr>
<td>ome om</td>
<td>3875</td>
</tr>
<tr>
<td>omu om</td>
<td>508</td>
</tr>
<tr>
<td>om ome</td>
<td>312</td>
</tr>
<tr>
<td>om omu</td>
<td>0</td>
</tr>
<tr>
<td>ome omu</td>
<td>15</td>
</tr>
<tr>
<td>omu ome</td>
<td>15</td>
</tr>
<tr>
<td>Column1</td>
<td></td>
</tr>
<tr>
<td>-------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>ome ome ome</td>
<td>16</td>
</tr>
<tr>
<td>omu omu omu</td>
<td>5</td>
</tr>
<tr>
<td>om ome om</td>
<td>13</td>
</tr>
<tr>
<td>om omu om</td>
<td>4</td>
</tr>
<tr>
<td>om om ome</td>
<td>10</td>
</tr>
<tr>
<td>om om omu</td>
<td>1</td>
</tr>
<tr>
<td>ome om om</td>
<td>179</td>
</tr>
<tr>
<td>omu om om</td>
<td>29</td>
</tr>
<tr>
<td>ome om ome</td>
<td>3</td>
</tr>
<tr>
<td>omu om omu</td>
<td>0</td>
</tr>
<tr>
<td>om ome ome</td>
<td>2</td>
</tr>
<tr>
<td>om omu omu</td>
<td>7</td>
</tr>
</tbody>
</table>

Figure 8 DL ALFAs for Three Consecutive Adjectives

<table>
<thead>
<tr>
<th>Series 1</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>og og</td>
<td>96972</td>
</tr>
<tr>
<td>oga oga</td>
<td>7598</td>
</tr>
<tr>
<td>og oga</td>
<td>11906</td>
</tr>
<tr>
<td>oga og</td>
<td>26239</td>
</tr>
</tbody>
</table>

Figure 9 G ALFAs for Two Consecutive Adjectives
<table>
<thead>
<tr>
<th></th>
<th>Series 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>og og og</td>
<td>4330</td>
</tr>
<tr>
<td>oga oga oga</td>
<td>263</td>
</tr>
<tr>
<td>og og oga</td>
<td>188</td>
</tr>
<tr>
<td>og oga og</td>
<td>311</td>
</tr>
<tr>
<td>oga oga</td>
<td>166</td>
</tr>
<tr>
<td>oga oga og</td>
<td>596</td>
</tr>
<tr>
<td>og oga oga</td>
<td>36</td>
</tr>
</tbody>
</table>

Figure 10 G ALFA for Three Consecutive Adjectives

<table>
<thead>
<tr>
<th></th>
<th>With Prep</th>
</tr>
</thead>
<tbody>
<tr>
<td>og</td>
<td>101716</td>
</tr>
<tr>
<td>oga</td>
<td>45353</td>
</tr>
<tr>
<td>om</td>
<td>345568</td>
</tr>
<tr>
<td>ome</td>
<td>76392</td>
</tr>
<tr>
<td>omu</td>
<td>7738</td>
</tr>
</tbody>
</table>

Figure 11 G ALFAs Preceded by a Preposition

<table>
<thead>
<tr>
<th></th>
<th>og</th>
<th>oga</th>
</tr>
</thead>
<tbody>
<tr>
<td>moj 'my'</td>
<td>307</td>
<td>198</td>
</tr>
<tr>
<td>tvoj 'your'</td>
<td>9</td>
<td>22</td>
</tr>
<tr>
<td>njegov 'his'</td>
<td>548</td>
<td>35</td>
</tr>
<tr>
<td>njezin 'her'</td>
<td>172</td>
<td>37</td>
</tr>
<tr>
<td>naš 'our'</td>
<td>689</td>
<td>212</td>
</tr>
<tr>
<td>vaš 'your (pl.)'</td>
<td>124</td>
<td>32</td>
</tr>
<tr>
<td>njihov 'their'</td>
<td>344</td>
<td>25</td>
</tr>
<tr>
<td>svoj 'one's'</td>
<td>2,885</td>
<td>1,575</td>
</tr>
</tbody>
</table>

Figure 12 G ALFA with Preceding Preposition for Possessives
Figure 13  G ALFA with Preceding Preposition for Demonstratives

<table>
<thead>
<tr>
<th></th>
<th>og</th>
<th>oga</th>
</tr>
</thead>
<tbody>
<tr>
<td>ovo 'this one'</td>
<td>1,758</td>
<td>1750</td>
</tr>
<tr>
<td>to 'that one'</td>
<td>4,473</td>
<td>17294</td>
</tr>
<tr>
<td>ono 'that one (over there)</td>
<td>896</td>
<td>2085</td>
</tr>
<tr>
<td>ovakav 'in this way'</td>
<td>82</td>
<td>17</td>
</tr>
<tr>
<td>takav 'in that way'</td>
<td>371</td>
<td>84</td>
</tr>
<tr>
<td>onakav 'in that way'</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>kakav 'in which way'</td>
<td>64</td>
<td>10</td>
</tr>
</tbody>
</table>

Figure 14  G ALFA with Preceding Preposition for other non-Descriptives

<table>
<thead>
<tr>
<th></th>
<th>og/eg</th>
<th>oga/ega</th>
</tr>
</thead>
<tbody>
<tr>
<td>koj 'which'</td>
<td>4592</td>
<td>2862</td>
</tr>
<tr>
<td>ko 'who'</td>
<td>13</td>
<td>731</td>
</tr>
<tr>
<td>prvi 'first'</td>
<td>2624</td>
<td>814</td>
</tr>
<tr>
<td>jedan 'one'</td>
<td>2712</td>
<td>578</td>
</tr>
</tbody>
</table>

Figure 15  G ALFA with Preceding Preposition for Descriptives

<table>
<thead>
<tr>
<th></th>
<th>og</th>
<th>oga</th>
</tr>
</thead>
<tbody>
<tr>
<td>hrvatski 'Croatian'</td>
<td>1018</td>
<td>274</td>
</tr>
<tr>
<td>veliki 'big'</td>
<td>680</td>
<td>104</td>
</tr>
<tr>
<td>glavni 'main'</td>
<td>500</td>
<td>101</td>
</tr>
<tr>
<td>novi 'new'</td>
<td>698</td>
<td>145</td>
</tr>
<tr>
<td>mali 'small'</td>
<td>159</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>om</td>
<td>ome</td>
</tr>
<tr>
<td>--------</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td><strong>moj</strong> 'my'</td>
<td>1,311</td>
<td>246</td>
</tr>
<tr>
<td><strong>tvoj</strong> 'your'</td>
<td>64</td>
<td>22</td>
</tr>
<tr>
<td><strong>njegov</strong> 'his'</td>
<td>1,790</td>
<td>35</td>
</tr>
<tr>
<td><strong>njezin</strong> 'her'</td>
<td>538</td>
<td>37</td>
</tr>
<tr>
<td><strong>naš</strong> 'our'</td>
<td>3,533</td>
<td>368</td>
</tr>
<tr>
<td><strong>vaš</strong> 'your (pl.)'</td>
<td>807</td>
<td>70</td>
</tr>
<tr>
<td><strong>njihov</strong> 'their'</td>
<td>1529</td>
<td>36</td>
</tr>
<tr>
<td><strong>svoj</strong> 'one's'</td>
<td>11,905</td>
<td>2,848</td>
</tr>
</tbody>
</table>

Figure 16 DL ALFA with Preceding Preposition for Possessives

<table>
<thead>
<tr>
<th></th>
<th>om</th>
<th>ome</th>
<th>omu</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ovo</strong> 'this one'</td>
<td>13984</td>
<td>2855</td>
<td>124</td>
</tr>
<tr>
<td><strong>to</strong> 'that one'</td>
<td>25529</td>
<td>42824</td>
<td>2471</td>
</tr>
<tr>
<td><strong>ono</strong> 'that one (over there)'</td>
<td>982</td>
<td>1239</td>
<td>66</td>
</tr>
<tr>
<td><strong>ovakav</strong> 'in this way'</td>
<td>230</td>
<td>23</td>
<td>2</td>
</tr>
<tr>
<td><strong>takav</strong> 'in that way'</td>
<td>1005</td>
<td>89</td>
<td>12</td>
</tr>
<tr>
<td><strong>onakav</strong> 'in that way'</td>
<td>19</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>kakav</strong> 'in which way'</td>
<td>912</td>
<td>53</td>
<td>9</td>
</tr>
</tbody>
</table>

Figure 17 G ALFA with Preceding Preposition for Demonstratives
<table>
<thead>
<tr>
<th></th>
<th>om</th>
<th>ome</th>
<th>omu</th>
</tr>
</thead>
<tbody>
<tr>
<td>koj 'which'</td>
<td>36555</td>
<td>1</td>
<td>10730</td>
</tr>
<tr>
<td>ko 'who'</td>
<td>515</td>
<td>1801</td>
<td>411</td>
</tr>
<tr>
<td>prvi 'first'</td>
<td>10217</td>
<td>1172</td>
<td>62</td>
</tr>
<tr>
<td>jedan 'one'</td>
<td>9874</td>
<td>898</td>
<td>88</td>
</tr>
</tbody>
</table>

**Figure 18** DL ALFA with Preceding Preposition for other non-Descriptives

<table>
<thead>
<tr>
<th></th>
<th>om</th>
<th>ome</th>
<th>omu</th>
</tr>
</thead>
<tbody>
<tr>
<td>hrvatski 'Croatian'</td>
<td>6727</td>
<td>1003</td>
<td>16</td>
</tr>
<tr>
<td>veliki 'big'</td>
<td>2121</td>
<td>95</td>
<td>7</td>
</tr>
<tr>
<td>glavni 'main'</td>
<td>1262</td>
<td>146</td>
<td>1</td>
</tr>
<tr>
<td>novi 'new'</td>
<td>3497</td>
<td>479</td>
<td>21</td>
</tr>
<tr>
<td>mali 'small'</td>
<td>844</td>
<td>82</td>
<td>6</td>
</tr>
</tbody>
</table>

**Figure 19** DL ALFA with Preceding Preposition for Descriptives

<table>
<thead>
<tr>
<th></th>
<th>Croatia Subtitled</th>
</tr>
</thead>
<tbody>
<tr>
<td>og</td>
<td>1972</td>
</tr>
<tr>
<td>oga</td>
<td>399</td>
</tr>
<tr>
<td>om</td>
<td>1678</td>
</tr>
<tr>
<td>ome</td>
<td>163</td>
</tr>
<tr>
<td>omu</td>
<td>10</td>
</tr>
</tbody>
</table>

**Figure 20** DL and G ALFA Distribution in Visual Media
APPENDIX F: DIALECT MAP OF BCS SPEECH
(Reprinted with permission from Robert Greenberg via his website: http://www.unc.edu/~rdgreenb)
LIST OF REFERENCES


Kašić, Bartol. 1604. *Institutiones linguae Illyricae (Osnove ilirskoga jezika)*. Rome.


Newmeyer, Frederick. 2005. ‘A reply to the Critiques of “Grammar is Grammar and Usage is Usage”’. Language, 81, 229-236.


Preston. Dennis. 1996. ‘Where the worst English is spoken.’ E. Schneider (ed.). Focus on the USA. Amsterdam: Benjamins, 297-360

Preston, Dennis and Daniel Long (eds). 2002. Handbook of perceptual dialectology, Volume II. Amsterdam: Benjamins


