Brigham Young University
BYU ScholarsArchive

All Theses and Dissertations

2015-06-01

Botheration and Recognition of Prescriptive Rules

Sara D. Smith
Brigham Young University - Provo

Follow this and additional works at: https://scholarsarchive.byu.edu/etd

Part of the Linguistics Commons

BYU ScholarsArchive Citation
Smith, Sara D., "Botheration and Recognition of Prescriptive Rules" (2015). *All Theses and Dissertations*. 5582.
https://scholarsarchive.byu.edu/etd/5582

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu.

Botheration and Recognition of Prescriptive Rules

Sara D. Smith

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Arts

Don William Chapman, Chair
Jesse Egbert
Alan Manning

Department of Linguistics and English Language

Brigham Young University

June 2015

Copyright © 2015 Sara D. Smith

All Rights Reserved

ABSTRACT

Botheration and Recognition of Prescriptive Rules

Sara D. Smith
Department of Linguistics and English Language, BYU
Master of Arts

Passions flare up around the use and "misuse" of prescriptive rules. Where there is variation in language use, language judgment usually follows—attaching value judgment to linguistic variants forms the foundation of prescriptive ideology in English. Prescriptive attitudes prevail among speakers and writers of English, who feel some pressure to use these forms to avoid a negative judgment.

This study surveyed American English speakers using Mechanical Turk to determine which types of rules—spelling, syntactic, morphological, and lexical—bother people the most and inspire the harshest judgments when violated. The surveys asked participants to identify a violated prescriptive rule in a sentence, found using the magazine and newspaper registers of the Corpus of Contemporary American English, and then to indicate how much they were bothered by the violation.

Results indicated that lexical rules separating subtle semantic differences—*i.e. farther* vs. *further*, *comprise* vs. *compose*—tend to be less bothersome and less recognizable than other types of rules. However, the type of category that a prescriptive rules falls under does not seem to explain why some rules are more bothersome or recognizable than others. It may be possible to generalize by assuming that lexical prescriptive rules will be less important to a general educated American audience than spelling or grammar rules, and that nonstandard dialectal forms will be even more bothersome. However, the ability to generalize these results is limited: there is some evidence for a "pet-peeve" effect. Individuals seem to simply be bothered by different rules, without strong patterns showing some types of rules sharply more important than others. Additionally other prescriptive rules, including those regarding *nauseous* and *dove* as the past tense of *dive*, were more recognizable and bothersome in their prescribed form than their proscribed, providing evidence for semantic shifts.

Keywords: prescriptivism, descriptivism, botheration, language judgment, standard English

ACKNOWLEDGEMENTS

This thesis has been a marathon, and just as in the marathon I ran the summer before completing it, I have been cheered and helped and loved at the sidelines and from afar, beginning to end.

I am especially grateful for the patience and encouragement of my thesis committee and for each member's dedication to the success of one student. Dr. Chapman provided the idea and helped it grow by reading sloppy drafts and teaching me one-on-one in his office. The statistics in this thesis owe their existence to Dr. Egbert, who spent hours patiently helping me run tests and interpret results. I could not have done this without his help. I'm also grateful to LoriAnne Spear for answering my questions and keeping me on track through the years.

My dad was the first person I knew with a graduate degree—a copy of his giant dissertation is still in my bedroom. His and my mom's support helped me find the courage to seek my own graduate degree and gave me the strength to finish it, even when I thought it was impossible. Thanks for listening to me whine, and for telling me to get back to work.

Friends, coworkers, fellow students, and roommates—too many to name, I hope they know who they are—provided welcome distractions and necessary cheerleading. Thanks guys for laughing at the word *botheration*, for taking so many practice surveys, for helping me make them, for studying with me, for telling me to stop reading and start writing, for asking about my progress, for attending my defense—for following and cheering every step as I started and finally finished this marathon of a thesis.

TABLE OF CONTENTS

LIST OF TABLES vii

LIST OF FIGURES viii

CHAPTER 1: INTRODUCTION 1

The Prescriptive/Descriptive Binary 1

The Place of Prescriptivism in Linguistic Study 4

Linguistic Approaches to Prescriptivism 7

CHAPTER 2: REVIEW OF LITERATURE 10

Standard English and Value Judgments 11
Ideology of Standardization 11
Value Judgments 12
Eighteenth-Century Roots 14

Value Judgment and Modern Notions of Correctness 15
Golden Age Myth 15
Economic Benefits 17
Social Indexing 19
Avoiding Judgment 21

Botheration Studies 25
Problems in the Study of Error 31
A Hierarchy of Error? 34

A Space for More Research 35

CHAPTER 3: METHODOLOGY 37

Research Questions 37

Study Design 37

Survey Design 38
Selecting Test Items 38
Selecting Test Sentences 43
Survey Instrument 45

Participants 47
Mechanical Turk 47

Participant Demographics 48

Analysis of Results 51
Preparing the Data 51
Botheration 51
Recognition 53

CHAPTER 4: RESULTS 55

Botheration 55
Category 55
Item 61
The Effect of Item and Sentence 69
Demographics 70

Recognition 72
Category 73
Item 76
Demographics 83

CHAPTER 5: DISCUSSION AND CONCLUSION 85

Research Questions: The Effect of Formal Category 85

The "Pet Peeve" Effect 89

Individual Prescriptive Rules 92
Dialectal Items 93
Prescribed vs. Proscribed 93
Pop-Culture Targets 96
Relative Pronouns 97
Rule Pairs 99

Demographics 100

Improvements upon Previous Studies 100
Test Items and Sentences 101
Error Identification 101
Participants 102
Prescribed and Proscribed Forms 102

Limitations 103

Future Studies 105

Conclusion 108

REFERENCES 112

APPENDIXES 116

Appendix A
Usage Guides Consulted in Selecting Test Items 116

Appendix B
Test Items 117

Appendix C
Excerpt from a Sample Survey Instrument 120

List of Tables

Table 1: Items with Fewer than 10 Botheration Scores 52

Table 2: Mean Botheration Scores for All Sentences by Category 56

Table 3: Mean Botheration Scores by Category for Prescribed Sentences 58

Table 4: Mean Botheration Scores by Category for Proscribed Sentences 59

Table 5:Ranking of Item Mean Botheration Score (All Sentences) 62

Table 6: Ranking of Mean Botheration Scores of All Items by Category 63

Table 7: Ranking of Prescribed Items by Mean Botheration Score 65

Table 8: Ranking of Proscribed Items by Mean Botheration Score 67

Table 9: Ranking of Proscribed Items According to Category 68

Table 10: Mean Botheration by Education Level 71

Table 11: Correct Error Identification by Category 73

Table 12: Crosstabs for Error Recognition by Category 74

Table 13: Error Recognition Crosstabs for Prescribed and Proscribed Sentences by Category 75

Table 14: Error Recognition Ranked by Correct Identification of Proscribed Sentences by Category 77

Table 15: Error Recognition Ranked by Correct Identification of Proscribed Sentences 79

Table 16: Items with Fewer than 10 Botheration Scores 90

Table 17: Prescribed Items with Higher Botheration Scores than Proscribed 94

Table 18: Prescribed Items Recognized as Errors at Higher Rates than Proscribed Items 95

Table 19: Comparison of Rule Pairs 99

Table 20: Results Summary Chart 109

List of Figures

Figure 1: Study Design 38

Figure 2: Participant Gender 49

Figure 3: Participant Age 50

Figure 4: Participant Education 50

Figure 5: Mean Botheration Scores for All Sentences by Category with Tukey HSD Groupings 57

Figure 6: Mean Botheration Scores by Category for Prescribed Sentences 58

Figure 7: Mean Botheration Scores by Category for Proscribed Sentences with Tukey HSD Groups 59

Figure 8: Mean Botheration Scores for Prescribed and Proscribed Sentences by Category 60

Figure 9: First Mixed-Effects Model 69

Figure 10: Final Mixed-Effects Model 69

Figure 11: Correct Error Identification by Category 73

Figure 12: Correct Error Identifiction of Lexical Items 81

Figure 13: Correct Identification of Morphology Items 82

Figure 14: Correct Identification of Spelling Items 82

Figure 15: Correct Identification of Syntax Items 83

Chapter 1: Introduction

As a field, linguistics tends to shy away from the study of prescriptivism. Prescriptivism is a cultural and social force in language; the ideology of "correctness" is often the only lens and meta-language English speakers have to view and discuss their own native tongue. Linguists should (and many do) take another look at the study of prescriptivism. My thesis strives to take a descriptive, quantitative look at unseen attitudes shaping the way native English speakers judge their own language, specifically by asking whether some types of prescriptive rules matter to them more than others.

The Prescriptive/Descriptive Binary

The ideology of prescriptivism is defined by Battistella as the conviction that "traditional grammar rules are based on logic, reason, and truth independent of usage," and that language and society are endangered unless perceived "errors" in grammar are avoided or corrected (2005:48). In the prescriptive tradition language change is resisted, and where variant forms exist one is typically deemed better or more correct; codified "rules" appeal to tradition and are claimed to represent the preference of the educated. Breaking a rule therefore represents lack of education as well as disregard for authority and tradition.

This ideology has long been challenged in the field of linguistics. One of the first lessons every new linguistics student learns is that linguists study, describe, and celebrate actual usage; they do not dictate how a language should be used. *Descriptivism* is the term often used to contrast the work of linguists—documenting and theorizing—from the judgments of prescriptivists. Descriptivism emphasizes "change over stability, diversity over uniformity, usage over authority, and the spoken language over the written language" (Drake 1977:9). In this definition Drake creates a series of binaries, with the

second and disfavored element in each pair representing prescriptivism—the disfavored half of a prescriptive/descriptive binary. In this binary, the term *descriptive* when applied to linguistics means less an objective look at language and more an attitude of "anti-prescriptivism." Prescriptivism becomes "the threatening Other . . . a specter that haunts linguistics" (1995:5), says Deborah Cameron, author of *Verbal Hygiene* (1995), who admits feeling isolated from fellow linguists for her interest in studying prescriptive attitudes.

While linguists who study prescriptivism may feel like a minority within their field, the ideology prevails in the world at large: "many millions of people are involved with language education, where a prescriptive hierarchy is rarely absent, while students of descriptive linguistics number somewhere in the thousands" (Joseph 1987:17). The majority of English speakers subscribe to a prescriptive approach to language; it is the ideology they are taught in school and the primary way they think about and interact with their own language. "We live in the age of the mavens," adds Jack Lynch (2009:112), noting that Lynn Truss's 2003 bestselling punctuation guide *Eats, Shoots & Leaves* is only one in a "long list of unexpectedly successful books on language." These grammar "mavens," the decidedly prescriptive authors of numerous books on English usage and style, are not linguists but specialists, writers, and hobbyists, often with a limited knowledge of the history of English or its structures. While linguists claim expertise on how languages work, prescriptivists claim to know and teach how language should be used: "a turf war between two groups of putative language experts has broken out," with linguists, who "have had limited success in writing for a popular audience," and mavens, who have a "limited knowledge of the research of professional linguists," dividing

themselves along the prescriptive/descriptive binary (Lynch 2009:112). Bill Walsh, a well-written and widely-read champion of the prescriptivist camp, called linguists "learned spoilsports" who "pooh-pooh our fun" (2013:8)—while linguist David Crystal refers to such fun as "quackery" (2006:125).

Linguists' criticisms of prescriptivism began with denouncements of eighteenth-century grammarians like Robert Lowth and his contemporaries and their unscientific, value-judgment–driven approach to English grammar: "The exclusion of prescriptivism which linguists have practiced since the emergence of scientific linguistics is due precisely to the qualitative value judgments which constitute it . . . . The exclusion of value judgments is part of the definition of 'scientific'" (Joseph 1987:18). The schism was solidified by the mid-twentieth century with the Chomskyan revolution in linguistics. If "grammar" is innately programed, as Chomsky (1986) and subscribers like Pinker (1994) asserted, then native speakers are the only judges of whether a usage is correct or grammatical—speakers instinctively know the "rules" without the aid of a grammarian or handbook of usage. Grammar, argue these linguists, is "embedded in the biological structure of the human brain" (Gilsdorf and Leonard 2001:445). Native speakers don't need to be taught rules and cannot make mistakes; they speak their mother tongue perfectly, having been born with a natural capacity to formulate grammar rules with an "innate linguistic endowment, common to all humans" (Joseph 1987:27).

Within this scientific ideology there is no room for value judgments; any claims that a standard or prestige variety is linguistically superior to any other form is nonsense (Milroy and Milroy 1999:10): "Prejudice against lower-class dialects is not dissimilar to racial and sexual prejudice. We believe that it is highly undesirable and that it is our job

as linguists to work against ignorance about dialect differences and for greater dialect tolerance" (Andersson and Trudgill 1992:123). On purely linguistic grounds no usage is "better"; every form is just as good as another, and prescriptions are linguistically arbitrary. Joseph (1987) cited numerous studies conducted in the 1970s in which subjects were unable to determine which variation of an unfamiliar language was the standard form, suggesting no intrinsic superiority of the standard dialect. "Generally, it is the speakers' prestige or lack thereof that gets transferred to the dialect" (Joseph 1987:59). Similarly, Pinker argues that

> most of the prescriptive rules of the language mavens are bits of folklore that originated for screwball reasons several hundred years ago. . . . there is no need to use terms like 'bad grammar,' 'fractured syntax' and 'incorrect usage' when referring to rural, black and other nonstandard dialects. (1994:20, 26)

The standard forms codified by prescriptivists are not "natural" ways of speaking; "they are learned behavior of precisely the sort Chomsky has insisted normal language acquisition is not" (Joseph 1987:17).

The Place of Prescriptivism in Linguistic Study

Why, then, would a study of prescriptivism have a place in a linguistics thesis? Joseph noted that linguists err when they "assume that the study of something which is non-scientific cannot itself be scientific: that a descriptive account of prescriptivism is not possible" (1987:18). My study attempts to provide a descriptive, quantitative study of the prescriptive attitudes that permeate the culture of English speakers and influence the way they interact with their own native language.

Beal *et. al* note an irony in linguists' sometimes passionate disavowal of the study of prescriptivism:

> The very people that advocate an approach to language that is based on principles of dispassionate rational inquiry fall into the trap of emotionally responding to a significant body of linguistic material and phenomena, thus preventing themselves from coolly analyzing it, trying to uncover the general principles underlying it and explaining any variation and change that a description of the facts may reveal. (2008:15)

This passionate disavowal leads linguists themselves to exhibit surprisingly prescriptive behavior. Cameron (1995) notes that the title of Robert Hall's *Leave Your Language Alone* (republished, perhaps wisely, as *Linguistics and Your Language* in 1960) represents the attitude of many linguists toward "language mavens." That attitude is itself a prescription dictating that "language would be better off without the constant unwelcome attentions of its speakers" (Cameron 1995:3). Cameron also notes that despite linguists' desire to avoid attaching value judgments, their distaste for prescriptivism is itself "ideologically non-neutral"—a value judgment (22).

Prescriptivism is part of the cultural and social behavior that influences language use. Standard and prestige varieties, and attitudes toward them, have a place within the study of linguistics. Sociolinguist Joshua Fishman noted that his field includes the study of "behavior and attitude toward language" (1982:3) and "consciously organized behavior toward language" (1982:2)—efforts to standardize language and the prescriptivists' practice of codifying their language certainly qualify as conscious behavior toward language. Sociolinguistics seeks to uncover "not only the societal rules

or norms that explain . . . behavior toward language in speech communities, but it also . . . determines the symbolic value of language varieties for their speakers" (1982:6).

While linguists view prescriptive attitudes as nebulous, relativistic, subjective, and arbitrary, excluding them from linguistic study

> eventually turns into a kind of willed ignorance. Value judgments about linguistic variants furnish a rich and barely tapped source of data about human cognitive faculties, and about how the gap between language and cognition becomes translated into norms. In so far as consciousness is a linguistic construct, value judgments on language represent language action upon itself, between planes of human consciousness. (Joseph 1987:5)

Feelings toward language usage have powerful social force. "When institutionalized these attitudes . . . fundamentally affect the lives of members of the speech community" (Drake 1977:5).

By studying prescriptive attitudes, linguists can examine how speakers and writers think about and interact with their own language. Milroy and Milroy note that languages and dialects "are not appropriate phenomena for value judgments," yet "speakers of languages do attach values to particular words, grammatical structures, and speech-sounds. There is apparently a yawning gap between what linguists profess to think about language and what ordinary people assume in their daily use and observation of language" (1999:11). This gap exists because, frankly, language speakers are not linguists or scientists when it comes to how they talk and write; they are members of a society, conscious of how the linguistic forms they choose in a given situation will affect their standing and perception within that society. The "rules" of standard English should

be important to linguists in that the speakers of English recognize that they exist, feel some pressure to follow them, and to some extent attempt to follow them (Wardhaugh, 1999:99–100).

These pressures are real for language users. Lynch sums up the passions that often flare up around issues like serial commas and expressions like *more unique*:

> Challenge my most cherished beliefs about the place of humankind in God's creation, and while I may not agree with you, I'll fight to the death for your right to say it. But dangle a participle in my presence, and I'll consider you a subliterate cretin no longer worth listening to, a menace to decent society who should be removed from the gene pool before you do any more damage. (2009:22)

Speakers and writers generally want to follow the rules of "good English," and, as Lynch notes,

> too often readers respond to the mavens' passion more than their knowledge. There's a strange phenomenon, little commented on by people who study the language: the rules we learn as children often stick with us, no matter how absurd, long after we should know better. (2009:115)

Perhaps it is one role of linguists to comment, or comment more, on the passion with which native speakers respond to violations of prescriptive rules.

Linguistic Approaches to Prescriptivism

A descriptive study of prescriptivism as a social and cultural force influencing language use has a place within linguistics. Two authors suggest specific binary-breaking, middle-ground approaches to a linguistic study of prescriptivism. The first is Cameron's concept of "verbal hygiene," a term she uses to describe activities and practices showing "an urge

to improve or 'clean up' language," ranging from belonging to a spelling reform society to taking writing classes to mocking another's pronunciation (1995:9). This refusal to stop tampering with natural language use and change does not entail universal agreement on what the norms of language should be. Rather it ultimately represents a social attitude affecting how language is used and perceived by its speakers: the attitude that the speakers have some sort of total control over their language, despite modern linguistic theories arguing that speakers are "largely unable to act upon it themselves" (Cameron 1995:17). Cameron argues that a study of English usage should consider verbal hygiene, which "is as basic to the use of language as vowels are to its phonetic structure, and as deserving of serious study" (1995:18).

Second, Battistella tries to navigate the "battle" between descriptive and prescriptive attitudes by adapting what he terms a "realist" position toward language, ultimately recognizing that while language canons and traditions are artificial, they are also an inevitable feature of a speech community's interaction with its own language and of its response to variation. A standard language is not important because, as prescriptivist subscribers claim, it represents the best usage of educated speakers, but

> because it is a cultural touchstone of the social history of the English language. . . . A realistic view of language also means understanding the inevitability of grammatical norms, of etiquette, and of a tradition of public writing. None of these can be ignored or dismissed by those seeking full participation in commerce, culture, and civic life. But variation and innovation cannot be ignored either. (2005:21–22)

Recognizing that prescriptivism is an important influence in the usage and language interaction of English-language speakers, the present study approaches prescriptivism, and particularly value judgment in language, from a descriptive standpoint. This study aims to quantify which types of variations in the standard language are most recognizable and inspire the harshest judgments. It will contribute to discussions on judgments about rules, specifically attempting to find a way to quantify which rules, and which types of rules, “matter” the most to native English speakers.

Chapter 2: Review of Literature

This research begins with a simple question: of the thousands of prescriptive rules filling usage guides and preached in high school English classes, which are most important? Perhaps the question is not so simple, given that there are potentially many ways to measure the "importance" of a rule. A researcher could find out which rules appear most commonly in handbooks—which ones matter most to the prescriptivists and others who codify rules and sell their books to writers and students. Language professionals, educated scholars, English professors, and writers—those who are supposed to teach, safeguard, and use the variety of educated English codified in prescriptive rules could be consulted. Maybe the most "important" prescriptive rules are those that are most often followed, or perhaps most often broken. Or, these rules are the ones that are most noticeable and bothersome when they are broken—with the least important rules slipping by unnoticed when disregarded.

Researchers within the field of composition education have surveyed academics and business professionals to determine which rules, when violated, bothered their participants the most. These studies, termed "botheration" studies and discussed later in this chapter, relied on judgments to determine which rules were most "important." In the case of these studies, the researchers wanted to know how harshly certain rule violations were judged by an audience their English students cared about impressing. The present research also takes this approach of attempting to measure how much certain rules or certain types of rules are judged—or rather, how "bothersome" the violations are. These botheration studies provide initial evidence that not all prescriptive rules are equally valued, nor are all proscribed forms equally noticeable.

Native English speakers are no strangers to being judged or judging others for "bad grammar." The fear of being judged is an important component to the culture and ideology of prescriptivism—many usage books are sold on the premise that they will help writers and speakers avoid this judgment, and the Internet is filled with websites, social media groups, and memes collecting, condemning, and correcting "bad grammar." Value judgment on variations in standard English—something that permeates prescriptive attitudes, asserting that one variation of the same form is inherently better than another—shapes a culture in which native speakers care deeply about prescriptive rules.

Understanding the rule of value judgments provides a background for the attitudes that may be shaping results of my study and shows a more qualitative backdrop to my quantitative research. In this chapter I will explore these attitudes of language judgment as well as literature about error and responses to error. As part of this I will discuss the work of the botheration researchers, who also attempted to quantify which types of violations of prescriptive rules inspire the harshest judgments, as their work provides some evidence that judgments of proscribed forms vary by type of rule.

Please note that in this section and in the remainder of this thesis, I will refer to violations of prescriptive rules as "errors" only for the sake of brevity; by no means do I equate a proscribed form with a true language error, such as those that might be made by a non-native speaker.

Standard English and Value Judgments

Ideology of Standardization

Despite how often the term language "standard" is found in the literature, the concept is more complicated than one might guess. The word has been applied to language since the

early eighteenth century, when notable writers called for English to be cleaned up and for changes in the language to stop (Crystal 2006:35; Joseph 1987:5). This desire became an important force behind the creation of dictionaries and grammars in that century. Today standard languages are generally prestige varieties taught in schools and used in mass communications, high culture, and government. It is the "safest" variety, claims Fishman, one that "stands for the nation as a whole" and can be employed when a speaker wishes to avoid judgment or to be the most understandable to the audience—when a speaker "cannot know his diversified and numerous listeners" (Fisherman 1982:24).

Fisherman also notes that the creation of a standard is a common behavior toward language, a "characteristic societal treatment of language, given sufficient societal diversity and a need for symbolic elaboration" (Fisherman 1982:18–19). Human intervention is "essential in the creation of a standard language; while some consider standard English to be a 'variety' of English, it is not one that developed naturally (Joseph 1987:15). Milroy and Milroy stated that the standard is an ideology rather than a variety—an ideology that represents an unchanging, carefully guarded, dead language (1999:22). Creating and following a standard is the concern of language "gatekeepers" such as writers, teachers, editors, grammarians, or other professionals, and is maintained in the speech community through grammars, dictionaries, and style manuals. Prescriptive rules attempt to codify standard English based not on empirical usage data but rather according to perceived usage and appeals to tradition.

Value Judgments

The prescribed forms said to comprise the standard are primarily acquired through education, and are so claimed to represent the forms used by educated speakers. (Trudgill

1990:117; McArthur 1999:167; Battistella 2005:42; Walsh 2013:42). Walsh and other commentators may claim that the use of certain variations by educated speakers make these forms standard, but perhaps it is more accurate to say that by using the prescribed forms, speakers appear to be educated. While speakers may claim to value "correctness" in English, they may not always use, or even be aware of, many prescribed forms. People believe in the existence of a correct form, but they don't always or consistently use these forms in their own writing or speech (Milroy and Milroy 1999:25).

In contrast to the standard language are all forms that exist outside of it: the non-standard dialects. Language commentators have tended to lump "the rest of English" together under titles like "nonstandard, substandard, and deviant" (Joseph 1987:30). Algeo (1991) similarly argued that for a linguistic issue to be considered a "question of usage," three factors must be present: there must exist alternatives of use; languages users must be able to choose among them; and those same or other users must think that the choice means something (2–3). Speakers must attach some sort of extra-linguistic value to one alternative because it carries social information or has a social function.

Where there is variation in language, whether lexical, phonological, or syntactical, there is also evaluation and the application of a value judgment. (Andersson and Trudgill 1992:4; Joseph 1987:30). "Just as there exists no human speech community without variation, there is no speech community, literate or otherwise, whose members are not consciously sensitive to language quality in one form or another" (Joseph 1987:3). And these variations in use or quality are subject to value judgment and "assignments of prestige" (1987:30). In *Eloquence and Power,* Joseph outlines nine characteristics of a standard language, and one of these is that the standard is the variety that carries a

positive value judgment among its speakers (1987:6). Other varieties different from the standard tend to be judged more harshly and their speakers considered less educated. Standard forms are judged positively while non-standard judged bothersome at the least, morally depraved at the most. These value judgments are so grounded in the culture that they seem "to represent a natural rather than an arbitrary order" (Joseph 1987:17).

Eighteenth-Century Roots

Attaching value judgments to linguistic variants forms the foundation of prescriptive ideology in English. It was a desire to avoid social judgment that led, in part, to a market for seventeenth and eighteenth century grammars. England's Industrial Revolution brought prosperity and class mobility, creating a "newly self-conscious group of people who were no longer peasants but were still excluded from the traditional aristocracy" (Lynch 2009:38). This socially mobile group needed guidance to navigate the customs of polite society—leading to the publication of conduct manuals that included guides to speaking and writing "proper English."

These guides evolved into entire books just for teaching English grammar to English speakers by authors such as Lindley Murray, Joseph Priestly, Robert Lowth, and others—by some estimates, upwards of 70 English grammar guides were published in the eighteenth century (Percy 2004:153). The authors outlined grammar principles such as parts of speech based on Latin classifications, sometimes including spelling or punctuation guides, and informed readers of the "correct" or "elegant" grammatical forms: "the placing of the preposition before the relative, is more graceful, as well as more perspicacious; and agrees much better with the solemn and elevated style," wrote Bishop Lowth of sentence-ending prepositions (96). "*Either* relates to two persons or

things taken separately, and signifies the one or the other. To say 'either of the three,' is therefore improper," declared Murray (66). While many of these "rules" remain in the books today, some grammarians' logic may seem odd, such as Miege's take on verb agreement: "And, forasmuch as the first Person is counted better than the second, and this better than the third, where two or three several Persons go before a verb, the verb agrees with the best Person. For Exemple, *You and I were very much concerned at it; You and he are always together.*"

The primary consumers of these grammars were the socially mobile, along with teachers and schoolchildren—many of them the children of the Industrial Revolution's rising class. This group paid to study grammar based on the assumption that "to speak properly was to speak like the traditional aristocracy" (Lynch 2009:45). This attitude reinforced the role of grammar and language use in marking class distinctions (Battistella 2005:47).

Value Judgment and Modern Notions of Correctness

The modern concern with prescriptive rules seems to center around the idea that one variety of English can be intrinsically better or more desirable than another. The reasons why some writers and readers value "proper English" include a sense of morality and the "golden age" myth, economic benefits, the desire mark themselves as part of a social group, and attempts avoid the negative social judgment that comes from using proscribed forms.

Golden Age Myth

Some in today's "age of the mavens" view "correct" language as a moral issue, with language change only another example of social and cultural decay. Even those who may

condone straying from other traditional social/cultural norms still hold onto traditional English usage, barely batting an eye at sexual promiscuity or shorts worn in a formal setting, for example, while frowning on non-literal uses of *literally.* While variation in any language is natural, many fear, distrust, or try to suppress some variations and innovations, believing these "errors" appeared because "an original, correct form of the language has decayed to its present state, and that it is a cultural duty to restore it," wrote Joseph, also noting that for some, the "decline" of language "signals the decline of culture itself" (1987:8). Any language change is a decline or decay:

> Some people see the standard language as representing linguistic health and see variation as a metaphorical infection. The prescriptive approaches certain fixed rules as defining the standards of clarity, logic, precision, and discipline, as respecting authority and tradition. For prescriptivists, disobeying the rule or changing them indicates a disregard for these qualities. (Andersson and Trudgill 1992:18)

Those who subscribe to prescriptivist attitudes—whether or not they've ever heard that word—look longingly back to a "linguistic golden age," asserting that current usage is inferior compared to the pure and authentic language of the past. However, as Lynch points out, this golden age never existed: "as soon as people stopped complaining about the illiteracies of the past, they switched to complaining about the illiteracies of the present" (2009:24). While speakers today may look to past decades for correct usage, back in 1962 Dwight MacDonald wrote that disregard for "traditional English" is "debasing our language by rendering it less precise and thus less effective as literature and less efficient as communication" (quoted in Lynch 2009:19). Centuries before that

commentators were also concerned about language change and expressed their disgust with current usage. Thomas Sprat in 1667 was one of the first to explicitly equate language "decline" with moral decline. In Sprat's history of the Royal Society of London, he stated that language change was "symptomatic of more general social evils, especially wars and the decline of religious practice" (quoted in Crystal 2006:63).

The connection between morality and language dates back to the foundations of Western culture. "The conceptual distance between absolute abstracts of any sort and religious beliefs is small," Joseph (1987:163) wrote, pointing out that in Judeo-Christianity, the creation of the world is implemented by God's words, God's name itself is identified with the copular verb, and linguistic variation results from God's wrath at Babel. "Language, passed on to humankind, establishes our primacy in the hierarchy of earthly creation. . . Thus the cult of the standard languages, and the Golden Age myth which underlies it, are not incidental features of Western civilization but its very cornerstone" (1987:164–65). The Golden Age myth remains integral to beliefs about language within Western cultures.

Economic Benefits

Beyond moral motivations, most English users care about prescriptions because they associate standard language with economic success. Just as in eighteenth-century England, the concept of social mobility today in the United States underlies modern notions of "good" and "bad" English. To many Americans, "better" speech and writing indicates a better education and better employability, so using prescribed forms and avoiding proscribed can lead to a better situation. Battistella writes that "grammar and language are part of the cultural capital that individuals pursue in order to improve their

social and economic situation … the idea that speaking and writing a certain way is the ticket to a better life seems to be ingrained in the public consciousness" (2005:11–12). The attitude prevails that "the right choice of words may give you the job you want; the wrong choice may keep you out of work" (Andersson and Trudgill 1992:4).

Therefore, the standard language is an economically empowering social tool: "Today's progressive argument is that fluency in a standard language is a necessary skill in a modern industrial society—for career entry and later advancement" (Battistella 2005:61). Employers, teachers, and other authority figures expect standard language, considering it to be essential to efficient, concise, widespread communication. These authority figures presumably reward those who master the conventions of the codified standard English (this idea, along with studies testing this presumption, will be explored later). Prescriptivist Bill Walsh in part justifies his work by arguing that "what you wear is right up there with the way you speak and write in forming the impression you present to the world" (2013:59)—in job applications and written documents, "correct" language can indicate competence.

Another group of people who benefit economically from the prescriptive culture include those specialists whose profession depends on enforcing the standards—teachers, editors, publishers, etc.—and authors of usage guides (such as Fowler) and popular guides to "good grammar" and writing (such as *Eats, Shoots & Leaves* and the works of Bill Walsh). This represents a small self-interested group with a specialized, economic interest in guarding the standardized English. For "professional verbal hygienists" maintaining norms is a business, a job that must continue to prove itself profitable and useful to society. Other speakers look to them as some sort of authority in how the

language ought to be used. As authorities and "keepers" of standardized English, they gain prestige and social function by their knowledge and enforcement of these rules. Cameron says that it is in editors' interest to "edit copy with extreme thoroughness, both to display conscientiousness and to maximize the hours for which they would be paid" (1995:52). She argues that "from an editor's perspective, then, hyperstandardization has its advantages: it makes a thorough editing job a relatively long job, a source of financial as well as professional satisfaction" (1995:52).

In considering who benefits from the attention copy editors give to language, Wardhaugh says that "one obvious group of winners are copy editors" (1999:4). Everyday speakers' worries about "falling standards" in English usage and self-consciousness about their own usage "ensures that their skills will continue to be valued" (Cameron 1995:40). Professional writing, argues Cameron, is a significant influence enforcing prescriptive attitudes: "the existence of a standard for writing is dependent on the existence of a standard for professionally produced writing." Editors and other language norm enforcers, as a matter of economic self-interest, create and follow styles and standards, fostering an illusion that consistency and "correctness" in writing is normal (Milroy and Milroy 1999; Cameron 1995).

Social Indexing

Because of the positive associations with knowing and using standard forms as discussed above—education, intelligence, morality, etc.—some English speakers and writers place importance on prescriptive rules in order to associate themselves with these positive attributes. While English speakers may or may not have a working knowledge of even close to every "rule" codified in the large number of usage guides, they can often spout

out a few "pet peeves" in language use: dangling participles, issues of case (*whom* vs. who), apostrophe usage (*it's* vs. *its*), mistaken homophones (*there*, *they're*, *their*), and subtle semantic shifts (*aggravate* and *anxious*). They may or may not use every rule consistently, but they outspokenly subscribe to the importance of using these rules and are quick to note mistakes in others. Why would these self-made mavens care so much about trivialities of usage?

William Labov in the 1960s and sociolinguists since then have shown that linguistic variation is affected by or associated with various social factors, like class, race, education, gender, and so on. Similarly, attempts to speak and write according to standards laid down by prescriptivist tradition are also examples of social indexing. A conscious choice to use or prefer a prescribed variant over the proscribed—even when larger usage of the prescribed form may be decreasing, as in *who* vs. *whom* or even stranded prepositions—may be an example of social indexing. The "standard" is presented as the way educated people speak and write; speakers using standard variants index themselves as members of a prestigious, moral, educated group: "Many individuals in today's society also judge a person's social standing, character, and propensity for adherence to a series of mainstream values such as honesty, thrift, and hard work by his or her use of standard English" (Heath 1980:31, quoted in Gilsdorf and Leonard 2001:445).

English speakers may feel motivated, even obligated, to use "correct" forms—to avoid and condemn abbreviations in electronic media, to insist on the pronunciation of *nuclear*, and to bemoan the decay of language in the hands of the next generation (Learmond 2010; Baron and Ling 2006; Tagliamonte and Denis 2008; Thurlow 2006)—

in attempt to index themselves as part of prestigious group. The standard is not "native" to anyone; it requires study and practice, and so those who can appear to master it are given social standing and consideration: "the prescribers . . . may be concerned with the maintenance and spread of the standard, but they simultaneously establish and manifest their own social superiority" (Joseph 1987:17).

Avoiding Judgment

The final reason I will be discussing regarding why writers of English may subscribe to a prescriptive ideology is a sense of linguistic insecurity. In a world full of people quick to judge everything from pronunciation to word use to spelling, English users do not only want to promote a positive judgment; they want to avoid a negative one from peers, parents, teachers, employers, etc. People with this concern may readily agree that they have "bad grammar":

> They know there's some difference between *lay* and *lie*; they know that *shall* and *will* are different somehow; they know there's some rule about where to put *only* in a sentence—and yet they don't know what those rules are. They've been scolded for confusing *can* and *may* . . . . They therefore have convinced themselves they're not using their language correctly. The only relief most people find is in the thought that at least some people speak worse than they do. . . . Most people speak improperly; only a talented and educated few get it right. (Lynch 2009:14)

The idea of following a rule is deeply associated with polite and correct behavior, and communities judge their members based on conformity to these behaviors. "Some believe that nonstandard language reflects unclear and incorrect thinking or that it arises

from a lack of initiative" (Battistella 2005:12). People have a guilt complex about the language they use; "few Americans, even among the well-educated, are confident and assured of the essential aptness and correctness of their speech" (Marckward 1958:50, quoted in Drake 1977:3)—perhaps out of fear of being judged because they have not mastered the "rules" and cannot claim to fall among the "talented and educated few" that Lynch mentioned. This insecurity also arises from assuming that edited prose is normal and naturally occurring; people often refer to this type of language when they talk about "grammatical rules" or "proper English."

> We are also supposed to try to make our spoken language conform to these rules. This requirement leads many to say that they do not know their grammar . . . They may be aware that they do not follow these rules when they speak, having been told this often enough and made to feel uncomfortable about it. (Wardhaugh 1999:105)

When considering the vehemence and scorn with which some describe language error, the desire to avoid negative judgment is quite understandable. Joseph Williams noted that "the language some use to condemn linguistic error seems far more intense than the language they use to describe more consequential social errors" (1981:153). For example, Bill Walsh wrote that people who use the proscribed phrase "I could care less" say "'could care less' because they've heard it all their lives and they are parrots. That might be a little harsh, but let's just say they're closer to being parrots than they are to being semanticists" (2013:20). His attack here moves beyond the grammaticality or logic of the phrase to a judgment of the people who use it. He continues: "If you wear sweatpants in public, I might think you're a slob. If you make a habit of parroting illogical expressions,

I might think you're on the slow side, or at least not much of a critical thinker" (2013:23–24). With judgments like this, is it any wonder that writers and speakers alike feel self-conscious about their "grammar"? Similarly, an entry in *Merriam-Webster's Dictionary of English Usage* (1994) summarizes commentators' responses to the non-reflexive, non-emphatic use of *myself:*

> Two general statements can be made about what these critics say concerning *myself*: first, they do not like it, and second, they do not know why. An index to their uncertainty can be found in this list of descriptors that they have variously attached to the practice: snobbish, unstylish, self-indulgent, self-conscious, old-fashioned, timorous, colloquial, informal, formal, nonstandard, incorrect, mistaken, literary, and unacceptable in formal written English (647).

It is also worth noting, as did the editors of Merriam-Webster in this passage, that the logic behind prescriptive rules is confusing or nonexistent, drawing wholly on tradition or personal preference. This muddies the water for writers trying to avoid judgment, as they may not know what forms will draw harshest criticism or understand the reasoning behind it, making it perhaps more difficult to remember to avoid the infraction that could get them accused of "ignorance, even of laziness" (Wardhaugh 1999:102).

Merriam-Webster's (1994) is just one of hundreds of usage guides published to collect or declare prescriptions. Many of these usage guides are relatively transparent about their purpose in helping writers and speakers of English avoid censure and earn praise. Ebbitt and Ebbitt's purpose, stated in their preface, is to help "students write so that what they have to say will be understood, respected, even enjoyed by readers."

Shaw's *Errors in English* (1986) promises to help readers communicate clearly, effectively "freed from restraint and anxiety," focusing on resolving errors "which hinder communication and impede thought" and "cause trouble, chagrin, or embarrassment" (p. xi). Bill Walsh's popular *Elephants of Style* (2004) sells itself with the cover tagline "packed with wisdom and expertise to help any writer look good." Other modern usage guides from Fowler's (Burchfield 2004) to Truss (2003) boast similar sentiments, which also carry the implication that if the usage items are not followed, writers and even speakers run the risk of being judged uneducated. Walsh bluntly addresses readers with descriptive inclinations: "if you think it's arrogant to condemn a perfectly understandable bit of prose as 'wrong,' you have to answer one big question: *Do you want to look stupid*?" (xii) By following the standards outlined in these usage guides, writers can avoid judgment—and if they don't, writers open themselves up to criticism from the reader, who could be anyone from a teacher to a boss or potential employer. This is the ideology most usage guides use to sell themselves, to prove their worth in the market.

Wardhaugh (1999) argues that it is committing an infraction, any infraction, that draws a negative judgment, and the"particular linguistic point is inconsequential is of little concern; in fact, it is quite irrelevant" (102). However, to the prescriptivist are some "linguistic points" more important to learn and defend than others? It would make sense that writers wishing to avoid judgment or to mark themselves as part of an educated group would attempt to learn and use those rules whose infractions are the most noticeable and frowned upon.

Are all the rules in the canon of equal importance, or do some inspire harsher judgment? That is the question that inspired the present study, along with many that

preceded it. Given the importance of attitudes of judgment—and the sharp, passionate feelings behind them—these attitudes and their potential to shape language are worth studying and measuring.

Botheration Studies

One group of researchers that attempted to measure attitudes of judgment toward individual prescriptive rules were what I will call the botheration researchers—a group of primarily composition professors starting with Hairston (1981) who asked the question, Which of the hundreds of rules are worth teaching in a college English class?

These researchers believed that with the growing number of usage guides and writers' handbooks filled with varying advice and prescriptions, writers, editors, students, and teachers—the main consumers and perpetuators of prescriptive rules—would ask themselves which prescriptions are most important to follow. Learning every rule of the thousands that have been codified in one book or another seems impossible and inefficient. There is no possible way that even language professionals could learn even the majority of these rules, much less the average person trying to impress a future employer. It follows that language users would want to learn only the "most important" rules—the rules that, when broken, inspire the harshest judgment.

The botheration researchers published primarily in 1980s, 1990s, and early 2000s, using surveys that measure "botheration," or how much an audience is bothered by a proscribed form, to ascertain the rules that would be most useful for their students to learn. "Standing in the language's midstream, we as teachers are responsible for teaching written English that is correct for our time," wrote composition teachers Gilsdorf and Leonard (2001:440). "It would seem futile to try to persuade students that the Standard

English of, say, 1930 (which some outdated handbooks still purvey) is still the Standard English expected by present and future employers." These researchers generally defined the "most important" rules as those that are currently most irritating or bothersome—something they weren't able to judge based on intuition or by usage guides and style manuals alone.

In the first of these studies, Hairston (1981) states the question behind her research succinctly: "What should our priorities be?" (794). Several authors were also motivated by accusations "from executives in business and technical organizations . . . that new graduates cannot write well" (Gilsdorf and Leonard 1990:141), and so attempted to find out what bothers employers most about new graduate writing, how that differs from what writing teachers think is important (Gilsdorf and Leonard 1990; Leonard and Gilsdorf 2001), and how they can better prepare students for expectations in the working world: teachers' "effectiveness . . . can be impeded if we stress matters that other professionals see as trivial—or if we trivialize points they deem consequential" (Beason 2001:34).

One of the difficulties inherent in these types of studies is the attention span and motivation of the participants. Hairston noted that her questionnaire contained 65 sentences: "fewer than I would have liked to use, but as many as I thought my readers would be willing to read and respond to" (1981:795). Leonard and Gilsdorf used only 45 items in 1990 (omitting dialect markers used in Hairston such as "he brung") and 50 in 2001. Young (1991) also used 50, and Kantz and Yates (1994) used 78; Gray and Heuser's (2003) update of Hairston's study used 88. Beason (2001) included only 20 items, though his qualitative study called for fewer subjects and test items.

These botheration studies, like the present study, attempt to rank errors and study reader reactions to proscribed forms. The researchers' interests were purely practical—they wanted to know which rules to emphasize based on the preferences of their students' intended audience: their college professors (Young 1991; Kantz and Yates 1994), future employers (Hairston 1981; Beason 2001; Gray and Heuser 2003), or both (Leonard and Gilsdorf 1990; Gilsdorf and Leonard 2001). In these studies, sentences with errors such as fragments, dangling modifiers, misplaced commas, and misspelled or misused words were given to business professionals or academics. The participants were asked to rate how much they were bothered by each sentence, assuming that "where readers are bothered by elements of what they are reading, they are likely to make some adverse judgment about the content, the writer, or both" (Leonard and Gilsdorf 1990:140). Each study showed significant variation in how "bothersome" each individual error was to its reader. Some researchers attempted to classify their results into types of errors, and though they neglect formal linguistic categories like syntax and morphology, the studies offer initial evidence that botheration may vary by type of error.

Hairston (1981) created a 65-sentence questionnaire, with each sentence containing a different "error," and asked responders to mark "does not bother me," "bothers me a little," or "bothers me a lot." Results showed that the most offensive proscribed forms were what she termed "status markers" such as *he brung* and *we was*. Her "status markers" are essentially dialectal features, most of which involve inflectional morphology. These forms are different from the others included in Hairston's and other similar studies as they are features of non-standard English, not simply proscribed forms that sometimes appear in standard English. These "status markers" were followed by

"mechanical mistakes" which included run-on/fragmented sentences and agreement issues, with punctuation and some lexical issues being the least bothersome. Hairston's study design did not satisfactorily isolate the desired proscribed form. Each item was tested only once, which led to the possibility of participants identifying the wrong error. Also, the use of constructed sentences doesn't remove the possibility that respondents were bothered by the unnaturalness or oddness of a particular sentence rather than solely by the proscribed form.

When Gray and Heuser (2003) replicated Hairston's study, results indicated a trend toward errors becoming less bothersome: number of "bothers me a lot" responses went down, while the "bothers me a little" responses increased. However, the hierarchy remained the same, with "status markers" being the most problematic. It's likely that "status markers" were so bothersome in these studies because they are dialectal items that are rarely considered an option in standard written English. Indeed, the other botheration researchers threw these items out, wanting to the include errors that their students would be more likely to struggle with. Gray and Heuser also improved on Hairston's design by adding two sentences for every error type, along with "correct" sentences and a "no error" response option. They found that the "no error" option was helpful, showing how many respondents missed an error and encouraging them to freely admit that they don't see an error on a test sentence, hopefully without fear of judgment. They also found that both sentences representing the same error rarely had the same botheration scores, showing the influence of factors beyond the error itself: not all errors are created equal, and neither are all contexts.

Beason tested this idea in his qualitative study that included a botheration survey followed by an interview with participants regarding why they made the choices they did. He also found inconsistent responses to the same errors, and interview evidence suggests "error gravity is easily mitigated or exacerbated by unique textual features that create or surround each instance of error—linguistic variables such as word choice, syntax, punctuation, or the location of these variables within the text" (2001:46). These results suggest the importance of including several examples of the same error within a survey, maybe even more than the four that Beason used. Beason's interviews also indicated that errors indeed harm the writers' image, as is claimed by the authors and marketers of usage guides, perhaps to an extent "more serious and far-reaching than many students and teachers might realize" (2001:48). Participants labeled writers as hasty, careless, uninformed, faulty thinkers, poor businesspeople, etc. Quantitative data showed some hierarchy to the error types, with fragments and misspellings the most serious and punctuation the least, though, as noted, botheration ratings varied within each example of the error.

Leonard and Gilsdorf (1990) also built on Hairston's study, this time testing not just professionals but academics as well. They used Hairston's scale and general design, but improved it by avoiding a convenience sample, including (and comparing) both professors and professionals, adding multiple examples of several errors to control for sentence awkwardness, and throwing out the dialectal items. Though time had passed, results were similar to Hairston's, with lexical issues (data as a plural, anxious/eager, etc) being the least bothersome, with sentence errors (run-ons, fragments) being the most. These authors updated their own study 10 years later (Gilsdorf and Leonard 2001),

dropping the least bothersome items from the survey and adjusting Hairston's three-point scale to a Likert. Again, they found little change in botheration: "interestingly, the 10 most distracting items in the 1990 study were nearly the same" (2001:49), with more variation in the least distracting items (likely because these were different in the two studies). In both 1990 and 2001, academics spotted more errors than business professionals and were more bothered.

Kantz and Yates (1994) used a more sophisticated design and analysis—one that helped inform the design of my study. They used real sentences as test items (while other researchers constructed them), taken from student papers and adjusted to include only one potential error. Overall, Kantz and Yates found evidence for a "hierarchy of rules" along with a "high level agreement across disciplines as to which errors were more or less serious." They included six sentences with no errors, and two to three different sentences testing the same error. They also asked participants (instructors and professors in a variety of disciplines) to mark the sentence error they responded to, and provided an option for "no error." This proved wise as it allowed the researchers to note that some errors were more recognizable than others. Those that bothered respondents the most were most often identified correctly as errors. Interestingly, "no teacher correctly diagnosed each error in the survey . . . every participant reacted at least once to some facet of a sentence other than the specific error that got the sentence chosen for the survey in the first place." This result points out a flaw in other botheration studies: if participants aren't given a chance to mark what part of the sentence they are responding to, or if the error itself isn't marked (as in Young's second section), the measurement of botheration for a specific rule is compromised.

Problems in the Study of Error

While these studies are useful—they offer initial results and models that have been helpful in the design of the current study—there remains room for improvement. One issue, as noted above, is the difficulty of isolating the intended prescribed form: "The results of our study, and of all similar studies, are weakened by the impossibility of researchers knowing for sure which part of a sentence participants are judging" (Gray and Heuser 2003:61). Another problem is that attention span leads to difficulties in testing large numbers of errors, especially if a researcher wants to include several examples of each rule used correctly and incorrectly to help isolate each individually. Another problem is the use of constructed sentences—an unnatural sentence may feel wrong to a reader simply because it is constructed. It would be better to replicate these studies using natural, attested sentences with errors or adjusted to include errors. The artificiality of the testing environment is definitely a drawback, though one that is difficult to resolve.

The audience surveyed in these studies is also important to consider—the participants were given credence as "experts" because they were teachers and employers, the same real-world audience held as the judges of college-student and college-graduate writing. But these groups are certainly not the only audience who matters. Armed with the rules they picked up in school, writers of English also have to face potential judgment from acquaintances on social media, coworkers and clients over e-mail, readers of a community newsletter, even potential significant others on dating sites. Newspapers and magazines are subject to the scrutiny of readers who may find any error to be a point against the credibility and quality of the publication. Prescriptions claim to preserve the language of the "educated"—which of course includes more than professors and

employers, such as doctors, judges, lawyers, government officials, etc. It might be of interest to linguists to explore which proscribed forms draw the highest judgment from various audiences, especially those considered to be the "educated" who regularly employ the language "standard."

Joseph William's paper "Phenomenology of Error" points out weaknesses in any study of error in language: "We are not always our own best informants about our habits of speech. Indeed, we are likely to give answers that misrepresent our talking and writing, usually in the direction of more rather than less conservative values" (1981:154). Thus when panels are consulted by dictionaries or usage guides about the use of the verb *impact*, "merely by being asked, it becomes manifest to them that they have been invested with an institutional responsibility that will require them to judge usage by the standards they think they are supposed to uphold." Whenever we ask for an opinion about a usage item, we need to be skeptical of our results; participants will likely see more errors than exist, and definitely more than they would in a non-artificial environment. Studying speakers' opinions is a tricky endeavor and rarely gives any real information about actual usage. The mismatch between usage and what people claim to think is a sociolinguistic paradox: "it seems to be virtually impossible to rely on speakers' reports of their own usage or of their attitudes to usage, so that we cannot easily find out what people actually think" (Labov 1966:214, quoted in Milroy and Milroy 1999:15).

Williams describes writing a previous manuscript in which he suggested that some errors are less serious than others (1981:155). His reviewers agreed with his suggestion, but each thought he was too soft on at least one violation and "unfortunately, each one of them mentioned a different item" (1981:155). As Williams illustrates, it is

important to be aware that botheration is idiosyncratic; linguistic "pet peeves" can vary by individual: "Great variation in our definition of error, great variation in our emotional investment in defining and condemning error, great variation in the perceived seriousness of individual errors. The categories of error all seem like they should be yes-no, but the feelings associated with the categories seem much more complex" (1981:155). Because of this idiosyncrasy, perhaps an analysis of botheration according to error type will be more effective, as it takes the focus away from individual "peeves" and attempts to generalize. Alternatively, an analysis of botheration could provide some evidence to support this "pet peeve" effect characterized by complicated feelings and experiences.

Williams also argues that when we read a document looking for errors, we will find them; and when we are not looking, we will not:

> If we read any text the way we read freshman essays, we will find many of the same kind of errors we routinely expect to find and therefore do find. But if we read those student essays unreflexively, if we could make the ordinary kind of contract with those texts that we make with other kinds of texts, then we could find many fewer errors. (1981:159)

He proves his point by sprinkling errors throughout his academic, peer-reviewed article—errors such as "the most obvioustest set of rules be those whose violation we instantly notes" (1981:160)—and confronts readers at the end of the article with the likelihood that they did not notice every error while reading as they were not on the hunt for error, given the formal context and their goal to engaged with the content rather than the text itself. The context of a published article does not prompt the search for error, while a survey asking participants to identify and respond to an error does exactly that.

A Hierarchy of Error?

Despite their weaknesses, the botheration studies do support the claim that not all usage rules are valued equally; some inspire harsh judgments and others seem unimportant. Though botheration results do suggest that participants likely have their own "pet peeves" items, the studies also provide initial evidence that variation does occur by type of error, not just error by error. However, there are likely other reasons why botheration varies. For example, Manning's review of Gilsdorf and Leonard's two studies argued that some errors are more bothersome than others because they impede sentence analysis and comprehension—the formal type of error matters less than what it does to clarity and meaning: issues of word choice may be disregarded, but what a structure does to clarity, whether or not a formal rule is involved, cannot be. "If a punctuation error is technically wrong but does not cause the misanalysis of the main-clause head, it will be less noticeable than any effort which does cause main-clause misanalysis" (2002:136).

Still, initial evidence does suggest that error category plays at least some sort of role in determining botheration: A meta-study combining data from several botheration studies provides enough evidence that botheration varies by formal category to suggest the usefulness of a larger quantitative study that investigates this issue. Chapman (unpublished) collected the botheration results of each item on the Hairston (1981), Young (1991), Kantz and Yates (1994), Leonoard and Gilsdorf (1990), and Gilsdorf and Leonard (2001) studies, and grouped each item into a formal category: punctuation, lexicon, morphology, spelling, and syntax. In order to compare results from different studies on the same scale, each item was assigned a *z* score—a standard statistical measurement that shows how far the botheration of that item is from the mean in that

study. The average *z*-scores for each group is different enough to suggest that the type of error category has an effect on how severe a judgment the error draws.

Chapman's results show that the morphology category draws the most judgment, followed by spelling, punctuation, syntax, and lexicon. "The ranking, as I will argue, is probably more consistent and therefore more useful for the high and low ends than for the middle. For a given item of a type at the high end, we may predict that the proscribed form would likely be bothersome. For a given item of a type at the low end, we may predict that the proscribed form would be less bothersome" (Chapman unpublished:24). It's worth noting again that the morphology category, especially in Hairston, often included dialectal issues like irregular verb forms that aren't viable options in standard English—perhaps this is why the morphology category drew such severe censure. For example, Johnson and VanBrackle's (2011) study of reactions to errors placed in student papers shows that "errors" representing African American English were judged more harshly than ESL errors and mistakes in standard English forms.

A Space for More Research

Given the importance and role of attitude of value judgment, the current study aims to quantify and measure these attitudes. They have been quantified in the past by botheration researchers, though not explicitly, by measuring attitudes of judgment toward specific prescriptive rules and proscribed forms. Evidence from these and other studies suggest that botheration may vary not just by individual rule but also by rule category, though no researchers have yet tested for this specifically, providing a space for the current research to examine which rules, or which types of rules, may be more "important" or meaningful than others.

If results confirm that botheration does vary by category, this study may provide some evidence for the possibility of generalizing—that is, if morphological rules on this survey are significantly more bothersome than other types, then perhaps any morphological rule that a teacher or other language professional runs across may receive greater priority than a semantic differentiation. It would also quantify which, if any rules, are more or less important than others.

Chapter 3: Methodology

Research Questions

Building off of the impetus question of whether some prescriptive rules "matter" more than others, the primary research question in this study is whether or not error botheration—how much people are bothered by the violation of a prescriptive rule—varies by the formal categories of morphology, lexical (semantic), syntax, and spelling. Secondarily, this study will question whether formal category accounts for variation in correct identification of violations of prescriptive rules. If botheration and/or recognition does not vary by category, the study hopes to provide evidence that some prescriptive rules are more important to English speakers than others, and perhaps identify some trends—or lack thereof—to describe which rules tend to be more bothersome or recognizable.

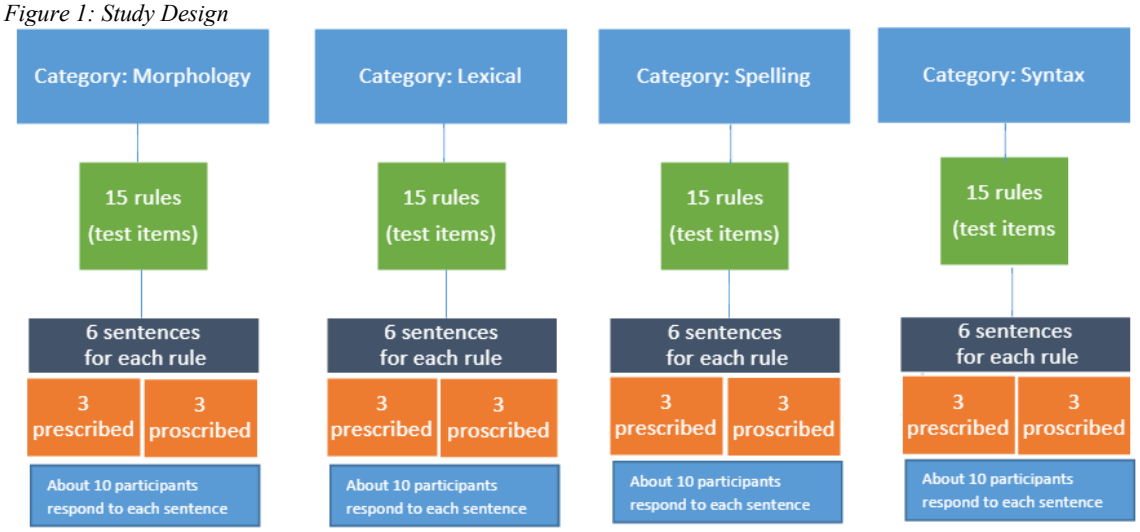
Study Design

This study uses data collected from surveys to measure the recognition and botheration of 60 usage items divided into four linguistic categories: morphology, lexicon, syntax, and spelling, with 15 items in each group. Each of the 15 rules, or test items, is represented with six different sentences, three with prescribed forms and three with proscribed. Overall, 60 items are tested using 360 total sentences. Figure 1 below illustrates the hierarchal structure of the study design, which will be explained in greater detail throughout this section.

The following section will outline how these 60 test items were chosen and how sentences containing published examples of both prescribed and proscribed forms for each item were found. Next I will describe the survey instrument and Amazon's

Mechanical Turk crowdsourcing website, which was used to recruit survey participants. The demographics of these participants will be described, followed with an outline of the methods used to analyze survey results.

Figure 1: Study Design



Survey Design

Selecting Test Items

The first step in creating a survey to measure botheration of proscribed forms is deciding which forms to test of the thousands of existing prescriptive rules. Most of the 60 prescriptions used in this study were drawn from among the rules most often addressed in usage guides published after 1970; the rest were chosen from items tested in other botheration studies.

As discussed in the literature review, thousands of prescriptions fill hundreds of usage guides, both popular and academic. For the purposes of this study, it is not necessary (or even feasible) to be comprehensive in testing reactions to every rule. This study attempts to discover whether there is any generalizing power in formal categories, so hypothetically, any rule within each category could be used. I attempted to select common, potentially more recognizable prescriptions to increase the likelihood that

participants would be able to identify and respond to the errors. Such "common prescriptions" would ideally be those that are currently most objectionable and perhaps most often confused—and most representative of their respective formal categories. Because quantitatively identifying the most well-known and well-recognized prescriptions could be a study and undertaking of its own, I elected to pull test items that appear most frequently in usage books published since 1970.

Usage manuals vary in which prescriptive rules they include—shorter, pithy guides marketed to college students and perhaps writers, such as Ebbitt and Ebbitt and Strunk and White, presumably have room only for items their editors deem most important or most often misused. The popularly marketed *Elephants of Style* (Walsh 2004) claims to address only the "gray areas" of usage—commonly debated items that need more discussion. Academic references like Fowler's (Burchfield 1996), Gardner (2003), or Merriam-Webster (1989) contain in hundreds of pages prescriptions both well-known and obscure. If a rule is included in only one guide or one type of guide, it is likely a less important prescription to observe, or maybe less known—otherwise, more authors would have picked it for inclusion. On the other hand, rules that appear in most guides, big and small, can be considered important in the prescriptive tradition.

Fortunately, over the past few years Chapman and his research assistants noted which prescriptions were contained in 35 usage handbooks published between 1926 and 2007. They found 13,000 items—some items are documented in only one or two guides, others appear in almost all of them. Chapman also collected the prescriptive rules tested in five botheration studies (Hairston 1981; Young 1991, Kantz and Yates 1994; Leonard and Gilsdorf 1990; Gilsdorf and Leonard 2001) and assigned each item a *z*-score to make

the results comparable, as discussed in the literature review. To choose items to test in the present study, I consulted both the list of rules compiled from handbooks and the list of items tested by previous researchers. My goal was to compile a list of items categorized evenly into four different groups: lexical, spelling, syntax, and morphology. I selected 15 from each category for a total of 60 items.

To select these test items, I consulted only items from usage handbooks in Chapman's list published since 1970 in an attempt to capture the rules that concern most recent writers and editors—though I do acknowledge that time, tradition, and precedent is extremely important. (A full list of the usage guides included can be found in Appendix A.) I manually sorted the top 100 most common rules on this list into one of the four formal categories.

I then used this list to begin selecting the most common prescriptions to include in the survey, considering each item individually. Some had to be disregarded. For example, the most common rule that fell within the morphology category was "a/an." This is an undisputed issue for which I was not interested in collecting data, so I excluded it. Another disregarded item was one of the top syntactic issues: dangling modifiers, a complicated issue with many facets and subcategories that could be the subject of its own study. I also consulted Chapman's list of items tested by other botheration researchers. From this list I added some of the most bothersome items, such as Hairston's "dialectal markers." Many of these "status markers" are variants of inflectional morphology found in non-standard dialects, such as *we brung* and *we was*. The later example is considered a morphological feature because unlike other instances of subject-verb agreement (about which there are many prescriptive rules), the question is not whether or not the subject is

plural, but how the verb is conjugated to reflect plurality. In some dialects, *was* is a plural verb form used to agree with plural subjects. The majority of these dialectal items fell into the category of inflectional morphology.

Though the most bothersome test items in previous studies, they didn't appear in the list of top prescriptions addressed in usage handbooks—yet all are addressed in at least one handbook. These items are not a part of standard English; careful writers and speakers don't think carefully or consult a usage guide when choosing between *we was* and *we were*. There is no argument based on usage or otherwise that would include these "status markers" as options or variations in standard English; they are features of nonstandard dialects and sound particularly odd to the ears of a speaker expecting standard English. This is likely why they were the most bothersome items in Hairston (1981). Though other researchers like Gilsdorf and Leonard elected to leave out the status markers, using instead "questionable usage elements" selected from "only those usage errors seen frequently in their business students' writing" (1990:142), I included some of these "status markers" among my test items, primarily to provide survey takers (and myself) a top-range measure of botheration. The items on my survey that fall in this category are *brang, we was, has went,* and multiple negation.

In considering both sources for test items, I also made sure to include several items that seem to be prevalent in popular culture—the kinds of rules cited in jokes about English majors, rules that people may have heard of and like to think they follow. In my survey these include the split infinitive, the stranded preposition, *who/whom*, and *literally*. Note that most of these fall into the syntax category.

In selecting items to test, in many cases I could choose to test a rule in one of two ways: I could test either the form or the function of the rule. In a discussion of empirical corpus research on usage, Chapman (unpublished) noted that some prescriptive rules are codified as "as a single function with competing variants"—a constant function with variable forms—while others are "formulated as a single form, with competing functions"—a constant form with variable functions (chapter 3:23). While some rules focus more on either form or function, Chapman notes that most rules could be construed either way. For example, for the word *disinterested* in the lexical category, I could choose to test the prescribed form of the word (*disinterested* with the function of meaning impartial) and the proscribed form of the word (*disinterested* with the function of meaning apathetic), rather picking one of these functions—the meaning of impartiality or the meaning of apathy—and testing the prescribed and proscribed forms. In this case, and most cases, I opted to test the rule for form. This was to provide consistency and also because it is often much easier to search the corpus for a form rather than a function. Appendix B, which contains a list of the 60 items selected, indicates whether or not form or function was tested.

Many of the items I selected in the spelling and lexical categories contained two possible forms to test. For example, the rule about *farther* vs. *further* can essentially be considered two separate rules: one governing the use of the form *farther*, and another governing *further*. The same goes for the rules *between* vs. *among* and the spellings of *all ready* and *already*; *discrete* and *discrete*; *awhile* and *a while*; and *effect* and *affect*. In all of these cases, I separated each form and tested it separately, so for the purpose of this survey, *farther* and *further* are considered different rules. There are other items on my

survey that could have been separated into two rules, including *comprise, disinterested, nauseous, adverse,* and *except.* In these and similar instances, I opted to test only one of the forms—*compose, uninterested, nauseated, averse,* and *accept* were not tested because these forms are less contested and less often used in their proscribed forms; for example, *nauseated* is rarely used to mean "causing nausea," but *nauseous* is often used to mean both "causing nausea" and "feeling sick."

Selecting Test Sentences

After selecting the prescriptive rules and deciding whether to test form or function, I turned to the Corpus of Contemporary American English (COCA) to find published sentences containing each of my test items. I searched item by item until I found six sentences for each: three with the prescribed form or function, and three with the proscribed form or function. Overall, I compiled six sentences for each of 60 items, for a total of 360 sentences.

Sentences primarily came from the magazine and newspaper registers, as these are published, respected sources with a hint of informality. They represent the kind of text with which an average educated American will interact on a regular basis—and the kind of text expected to be error free. At times I struggled to find a proscribed form in these two registers and so looked to spoken and academic for examples. Only rarely did I change a prescribed form in a sentence to a proscribed one.

While there are many advantages to using COCA to locate test sentences—e.g. they have been published, they are easily accessible, they are real samples of English usage—there are several drawbacks as well. Because the sentences come from published sources, they have likely been passed through one or more editors, who are paid to catch

and remove proscribed forms. Because of this, it was sometimes difficult to find a good number of examples of proscribed forms in the corpus; as Deborah Cameron says, "the marks of editing are all over published writing" (1995:34). Another challenge was isolating a sentence that can be understandable when taken out of its original context; several sentences required further editing, mostly to remove ambiguity, potentially confusing references to context, etc.

The benefit of using sentences in a corpus of published texts is to attempt a control for sentence awkwardness and intelligibility—the contrived sentences used in previous botheration studies may have had some influence in reported botheration, especially since the test item was not identified. As the test item is not identified in the present study, I took care to select sentences that can be intelligible outside of immediate context, and in some examples I clarified pronoun references or deleted clauses to simplify the sentences. Sentences that could potentially contain more than one instance of usage variation were skipped or edited. For example, one edited sentence contained the phrase "adverse effect," and because both *adverse* and *effect* are the subject of prescriptive rules.

It should be acknowledged that some test items had more obvious prescribed versions than proscribed. Joseph Williams noted that in approaching grammatical error and responses to it, there are two variables: "Has a rule been violated? And do we respond? Each of these variables has two conditions: A rule is violated or a rule is not violated. And to either of those variables, we respond, or we do not respond" (1981:159). Thus, a prescriptive rule may be used correctly (- violation), but even so may draw attention to itself and evoke a response (+ response), even if that response is a just a quick

internal acknowledgement that the rule was used correctly. For example, consider the following sentence from the present study using the prescribed form of *further*: "The ideal is to fall in love, then to fall further in love each passing year." A reader may pause to note or evaluate the correct use of *further*; participants in a study or testing environment may even think, "OK in this sentence they are testing my knowledge of the *further/farther* rule." Other prescribed forms, however, are less likely to be spotted. For example, the prescribed versions of test items *brang* (*brought*), *very unique* (*unique*), *we was* (*we were*), and split infinitives likely draw no attention to themselves.

Survey Instrument

It was not practical to use all 360 sentences on the same survey, so I created a series of 30 surveys, each testing 12 sentences. A sample surveys can be found in Appendix C. The surveys were created using Qualtrics survey software. The instrument and study was approved by the Institutional Review Board of Brigham Young University in December 2014.

The instrument first collected demographic information including native language (only native English speakers were allowed to complete the survey), age, gender, and education level. Participants were then presented with each of the sentences, and were told that each was taken from a newspaper or a magazine and may or may not contain an error (see exact wording in Appendix C). After they read one sentence, they were asked to identify whether or not they saw an error. If they reported seeing no error, they continued to the next sentence and were asked the same question. If they reported seeing the error, they were then asked how much the error bothered them on a 1 to 6 Likert-type scale, with 1 representing not bothered at all and 6 representing extremely bothered. Next

participants were asked to identify in a small text box the word or words containing the error—this response was used in data analysis to identify whether or not participants were responding to the target test item or to something else that bothered them in the sentence.

Each survey was pilot-tested among my acquaintances and Mechanical Turk (the marketplace used to recruit participants; to be discussed in the following section) to check for typos or errors beyond those being tested. Much of the feedback I received indicated that pilot participants—particularly those I knew personally—were afraid that *they* were the ones being judged on their knowledge of prescriptive rules and thus approached the instrument with anxiety and care. I had hoped that the participants would be the judges and look at the sentences critically without feeling too much pressure to get the answers "right;" I tried to adjust for this by making it explicit in the instructions (again, see Appendix C), but this fear of judgment is perhaps inherent in an instrument like this one.

To assign each of the 360 sentences to one of the 30 surveys, I assigned each sentence a random number, sorted the list by these random numbers, and then sorted according to the four categories. The first three from each category were assigned to the first survey, the second three from each to the second survey, and so on. Each survey contained an equal number of items from each of the four categories, but no effort was made to control how many prescribed and proscribed appeared in each. The likelihood that any survey would contain only prescribed sentences or only proscribed sentences was very small.

In every survey, in fact, participants discovered that they were almost as likely to encounter a prescribed form, a “correct” sentence, as an “incorrect” one. This is probably the biggest departure from the methodology employed by the other botheration researchers, most of whom used either no or very few “correct” sentences. Using prescribed forms keeps survey participants honest in their hunt for error. As Joseph Williams noted, the test-taking environment of a survey asking them to identify errors primes participants to see error, and may even encourage them to find error where none exists. Knowing that not every sentence contains an error, and seeing sentence after sentence with no error, may help mitigate the tendency noted by Williams. It also allows the researcher to equally measure responses to prescribed forms—are some prescribed forms, the “correct” forms, tripping up survey takers as much as, or even more than, the proscribed forms? Including an equal number of both in the study allows us to ask and perhaps answer these types of questions.

Participants

Mechanical Turk

Participants were recruited using Mechanical Turk (MTurk), an online marketplace hosted by Amazon.com for workers (“Turkers”) willing to complete menial tasks for small rewards. Requesters post small tasks, called Human Intelligence Tasks (HITs), which cannot be completed by a computer. These tasks range from tagging photos to verifying database information, and requesters generally pay a few cents ($0.05, for example) per task. Thus companies can cheaply crowd-source menial tasks instead of paying an employee to complete them on site. Requesters can review a Turker’s work and refuse to pay if the work was deemed of poor quality.

Recently, social researchers have seen MTurk as an opportunity to find an inexpensive, motivated, representative audience for research tasks (Mason and Suri 2011). Sprouse (2010) found results from a syntax survey administered in a lab and on MTurk to be comparable and almost indistinguishable. Perhaps results from MTurk can be considered better than those from a lab-administered survey, as it provides "easy access to a large, stable, and diverse subject pool" (Mason and Suri 2011:2) at a low cost, without recruiting only students or subjects connected to the university and the university's surrounding community. There are more than 100,000 workers on MTurk (4), and a three-year-long demographic study of the site by Mason and Suri found that 55% of workers are female and 45% male, with a median age of 30.

I hoped MTurk would inexpensively connect me with a general American audience in various locations. Because my survey includes several tasks and because I wanted to attract more participants quickly, I offered $1.00 per survey—a fairly high wage—in compensation. Funding was provided by the Linguistics and English Language Department of Brigham Young University. I requested 10 participants for each of the 30 surveys I placed on MTurk, and participants were allowed to take the same survey only once, though they could take multiple versions of the survey.

Participant Demographics

The survey was taken by 310 individual participants. Some surveys were completed by more than 10 people, and some survey sentences had too many confusing wording problems so I threw them out and retested the rules with different sentences, which resulted in additional participants. A total of 10 participants took multiple versions of the survey—one took it 7 times; the rest took it only twice.

Below, Figure 2, Figure 3, and Figure 4 represent the reported gender, age, and education information for the survey participants, respectively. As suggested by demographics from previous MTurk studies, gender for participants in the present study was split down the middle, with 52% of respondents reporting to be male and 48% of them female. The largest number of participants reported to be age 20 to 30 (41%), followed by age 31 to 40 (29%). The education results were fairly interesting: the highest percentage, 38.8%, reported completing four or more years of college, and 11.9% reported a graduate degree, with 22.3% reporting 2 years of college. In all, 81% of the participants reported having completed at least 1 year of college, showing that the respondents to this survey were fairly educated people. Though the study did not necessarily set out to measure the attitudes of the educated toward violations of prescriptive rules, the demographic results suggest that the results represent in part the reactions of fairly educated people.

Figure 2: Participant Gender

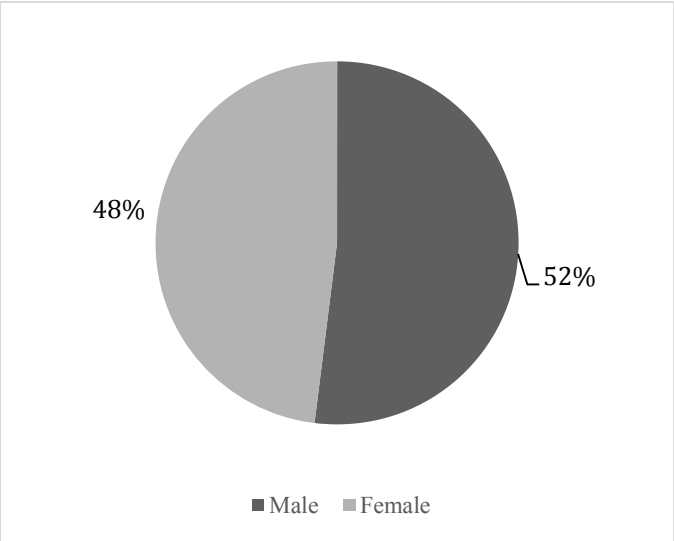


Figure 3: Participant Age

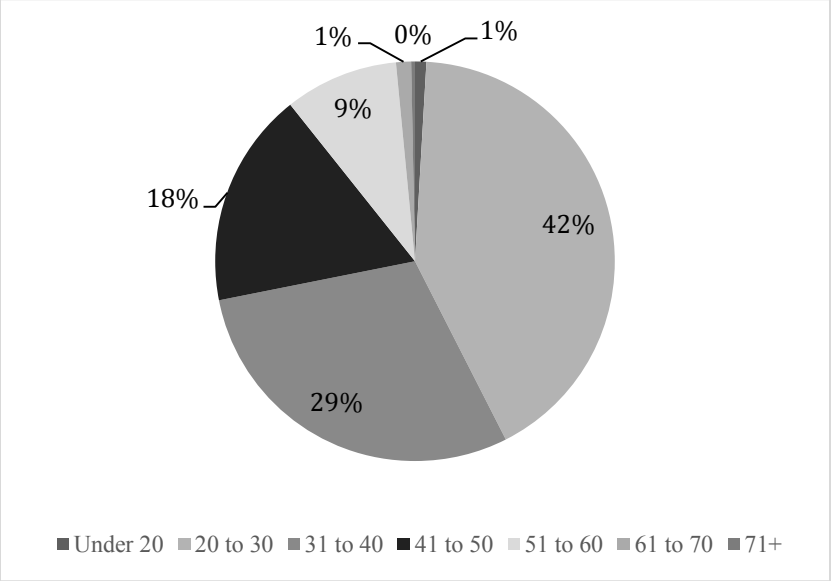
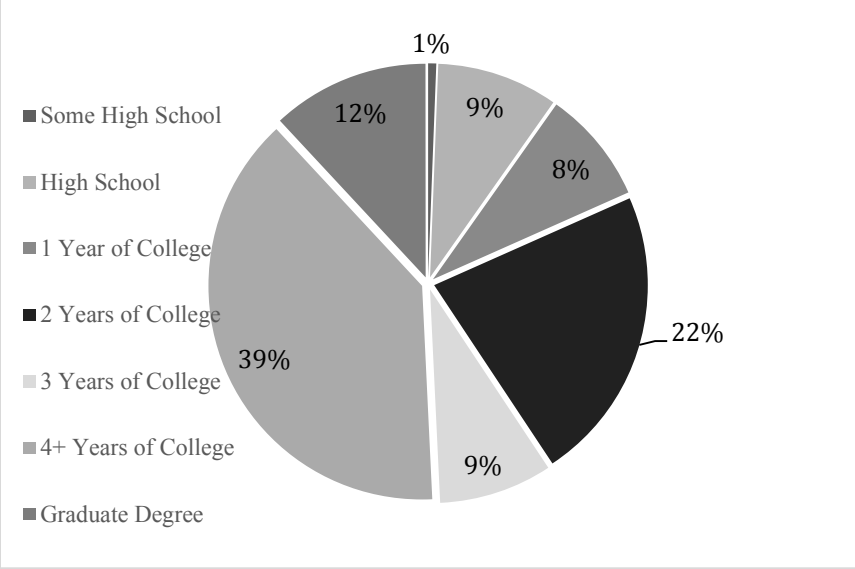


Figure 4: Participant Education



Analysis of Results

Preparing the Data

The data from each of the 30 completed surveys was compiled into a single dataset for analysis. Before these survey results were analyzed, each text response to the prompt "please type the word(s) that contain the error" was considered. If participants indicated that they observed an error on one of the sentences but typed in a response that did not contain the test item, their response to the question "Does this sentence contain an error?" was recoded from "yes" to "no," with the assumption that if they did not indicate seeing an error with the test item of interest, they did not see one. Therefore, the analyzed data has been cleaned up and recoded to ensure that only recognition and botheration scores for prescribed or proscribed test items was considered.

Botheration

The hierarchical nature of the study design—four categories comprising 15 items, each with six sentences, three of them containing prescribed forms and three with proscribed—lent itself to a mixed model analysis. This was completed using the statistical program R to explore how much variability in botheration ratings was accounted for by each of these four factors. Category and error present were set as fixed effects, and item and sentence as random effects, with sentence nested inside item.

Not every item received a botheration score—only those that were recognized by survey takers as having an error. Overall, 784 sentences (within 57 items) were recognized as having an error (and thus received a botheration score) Both prescribed and proscribed forms could be (and were) recognized to be errors. Before statistical analyses were performed to examine possible causes of variation in botheration scores, items that

had a botheration score but fewer than 10 respondents rating them were removed from the data. Overall 24 items were removed, leaving behind 36 of the original 60 for analysis. These removed items are listed below in Table 1.

Table 1: Items with Fewer than 10 Botheration Scores (N Lower than 10)

Item	Category	N	*M*
aggravate	Lexical	0	0
between	Lexical	0	0
split infinitive	Syntax	0	0
different than	Lexical	1	6.0
very unique	Lexical	1	2.0
disinterested	Lexical	2	3.0
adverse	Lexical	2	4.5
among	Lexical	3	4.0
principle	Spelling	3	4.67
that/which	Syntax	3	1.33
only	Syntax	3	2.67
who/that	Syntax	3	1.33
comprise	Lexical	4	2.5
discreet	Spelling	5	3.4
singular "they"	Syntax	5	2.4
stranded preposition	Syntax	5	2.8
Less	Lexical	6	2.17
hung/hanged	Morphology	6	3.38
discrete	Spelling	7	2.86
try and/try to	Lexical	8	3.0
who/whom	Syntax	8	3.38
reason is because	Syntax	8	3.0
infer	Lexical	9	3.56
all together	Spelling	9	2.44

Botheration was analyzed and listed according to category, item, and demographics (age, gender, and education level), and at each level all items were analyzed together and also with prescribed and proscribed sentence separated. Statistical tests were not performed at the sentence level because of the large number of sentences

and because results from the mixed-effects model analysis suggested further analysis would not be necessary. Statistical tests for botheration were performed using SPSS statistical software.

At the category level, an analysis of variance (ANOVA) was performed with Tukey *post hoc* tests to determine how the categories ranked in terms of mean botheration, if the difference between the groups was significant at the $p < .05$ level, and where those differences might be. At the item level, items were rank-ordered by mean botheration score to create a list of the most (and least) bothersome items overall and in each of the four categories. An ANOVA with Tukey *ad hoc* tests was also performed to compare the mean botheration scores of each item.

Statistical tests were also performed to measure the impact of demographics on botheration. For gender, an independent-samples *t* test was performed. For age and education, correlations were used, followed by univariate ANOVA tests with Tukey *ad hoc* tests.

Recognition

Results for recognition were analyzed using chi square tests, crosstabulations, and frequency counts. Error recognition was determined by the participants' response to the first question that appears with each sentence, asking if they see an error or not. Every sentence has an error recognition score, because this question was asked with each one. I was interested in finding out which categories and items were most often correctly (or incorrectly) identified as an error and thus used simple frequency counts to rank order error recognition.

An error identification was considered “correct” a. when a participant indicated seeing an error in a sentence containing a proscribed form, and then identified the test item in the text entry block; or b. when a participant indicated seeing no error in a sentence containing a prescribed form. An incorrect error identification occurred when a participant saw no error in a sentence with a proscribed form or identified another part of a proscribed sentence (other than the test item) as an error, or if a participant identified a prescribed test item as an error.

As with botheration, recognition was analyzed at the level of category and item, and at each level divided into prescribed forms, proscribed forms, and all forms. In the case of recognition, the analysis of proscribed forms proved to provide the most useful results. This is because, as noted above, a sentence with prescribed forms identified as having no error by a survey participant was considered a correct identification. There were many of these—it is perhaps more challenging to identify an error in a sentence then to read a correct sentence and recognize its correctness. Including the prescribed forms seemed to inflate the recognition results, making it appear that participants recognized “errors” at higher rates.

Chapter 4: Results

This chapter will present the data and results, beginning with botheration results and followed with error recognition. The results are presented in this order because botheration relates to the primary research question of whether botheration scores vary by formal category.

Botheration

Analysis of botheration scores attempted to account for variation among these scores—*i.e.*, to what extent do category, item, prescribed/proscribed form, sentence, and demographic information (age, gender, education) account for this variation? As explained in the previous chapter, variation was analyzed using a mixed-effects model, ANOVAs, and *post hoc* Tukey HSD tests. In the following section, the results for category and item ANOVA and *post hoc* tests first with all sentences will be presented and discussed, and then again with prescribed and proscribed sentences separated out. Finally, the results of a mixed-effects model testing all levels of the study will be presented, followed by data showing the effects of demographic factors on botheration scores.

Category

All Sentences

Means show that when botheration scores for all sentences are compared by category, spelling items (M = 3.76) and morphological items (M = 3.74) were the most bothersome with nearly identical means, followed by syntax (M = 3.58) and then lexical items (M = 3.19) (see Table 2). On the 1 to 6 Likert scale used in the survey instrument, these mean botheration scores all fall generally in the middle of the range, between 3 and 4. For

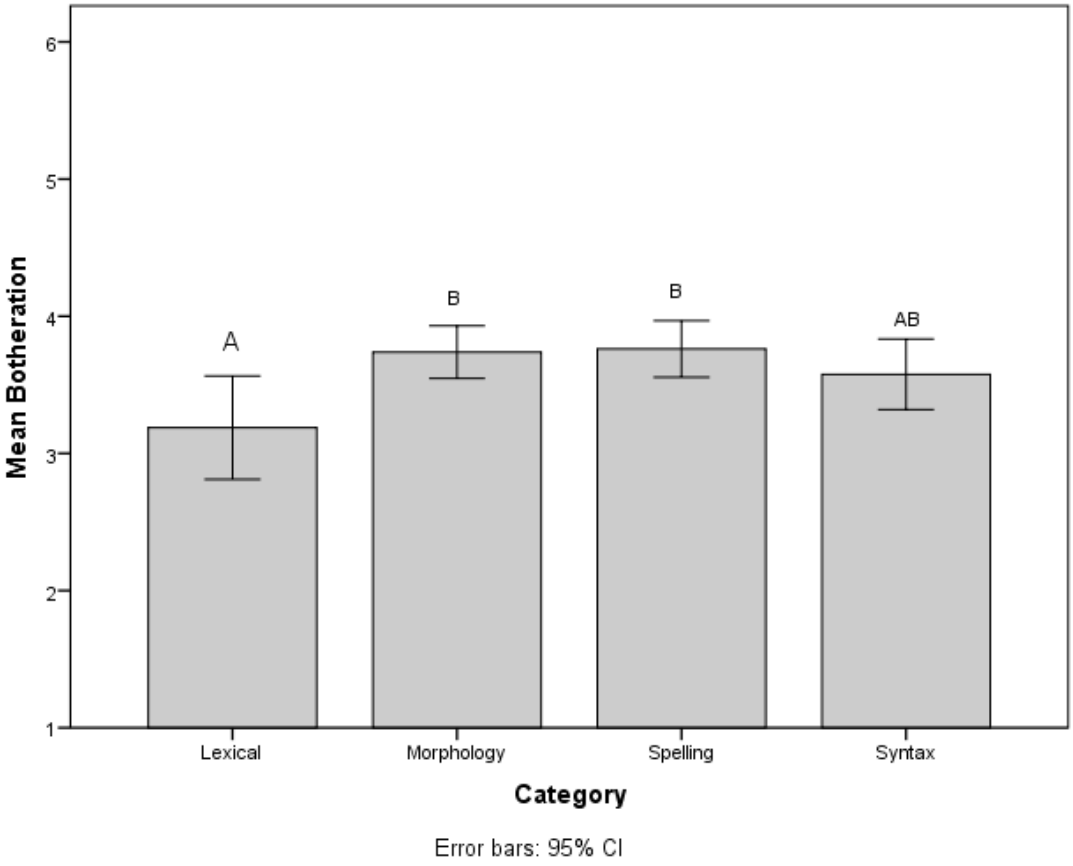
every category, there was a lot of variability, and standard deviations are relatively high. At least one item in each group was scored at 1 and one at 6. The botheration scores did group according to category, but the difference between the categories is not large.

Table 2: Mean Botheration Scores for All Sentences by Category

Category	*N*	*M*	*SD*
Lexical	64	3.19	1.511
Morphology	264	3.74	1.580
Spelling	218	3.76	1.541
Syntax	137	3.58	1.523
Total/Ave	**683**	**3.566687**	**1.538976**

A one-way ANOVA showed that there was a significant difference in botheration scores for all sentences by category, F (679, 1) = 2.651, p = 0.048, R^2 = .012. This low R^2 score shows that category accounts for a very small proportion of the variance in botheration scores—about 1% in this case. The Tukey HSD *post hoc* revealed that at the $p < .05$ level, lexical items received significantly lower botheration scores than morphological and spelling, but there was no significant difference between other groupings. The Tukey HSD groupings (there were two in this case: A and B) are identified with letters in Figure 5 below, which also displays botheration means and error bars. Members of the same groupings (i.e. groups that share the same letter) are statistically similar to each other, while those in separate groups are statistically different. Again, the chart illustrates that lexical items (group A) are the only group that is statistically different—lower, in this case—than other categories (group B).

Figure 5: Mean Botheration Scores for All Sentences by Category with Tukey HSD Groupings



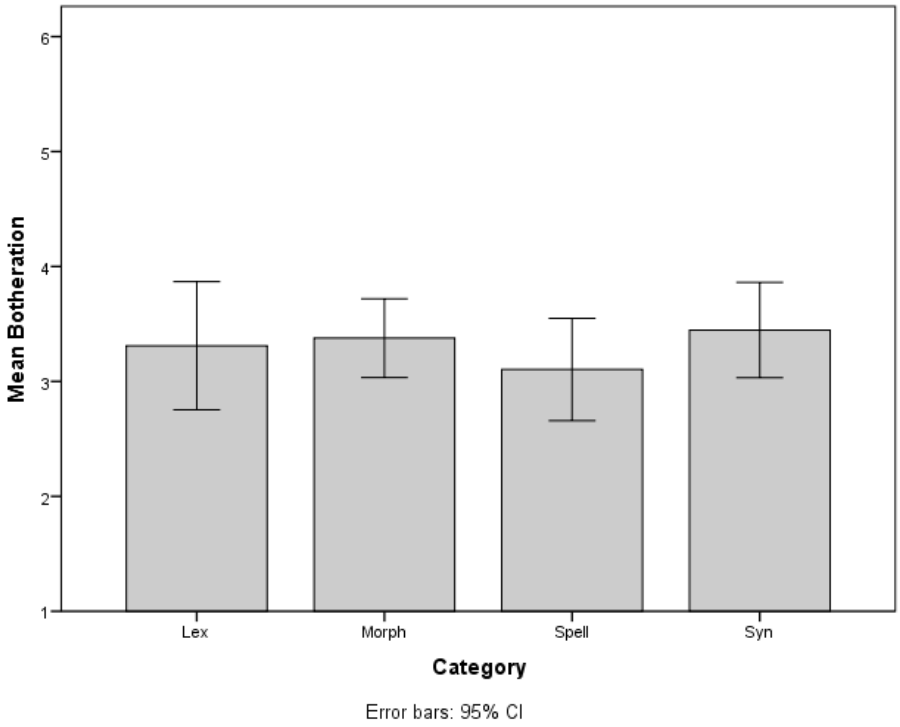
Prescribed Sentences

When looking at botheration by category for just prescribed sentences (sentences with no error in which the test item still prompted an error recognition and botheration score), morphological items prompted the highest mean botheration scores, followed by syntactic, lexical, and finally spelling. Table 3 and Figure 8 below show the mean botheration scores for prescribed sentences. A one-way ANOVA showed that there are no significant differences between any of these pairs of prescribed sentences (F [205, 1]= .471, p = .703).

Table 3: Mean Botheration Scores by Category for Prescribed Sentences

Category	*N*	*M*	*SD*
Lexical	29	3.31	1.466
Morphology	85	3.38	1.588
Spelling	48	3.10	1.533
Syntax	47	3.45	1.411
Total/Ave	**209**	**3.31**	**1.500**

Figure 6: Mean Botheration Scores by Category for Prescribed Sentences



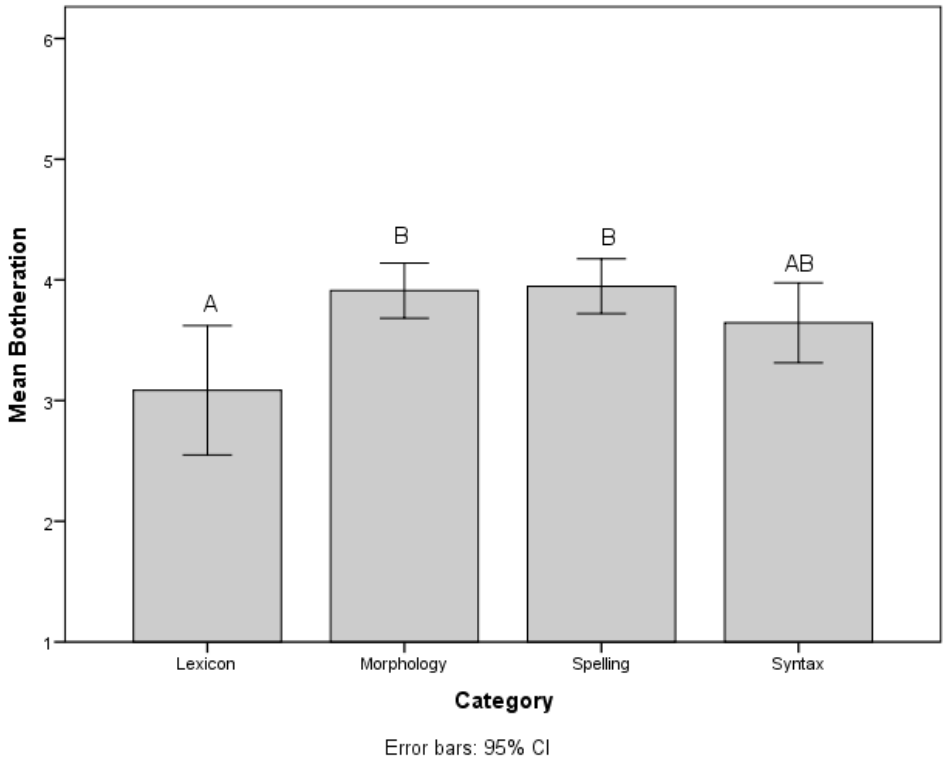
Proscribed Sentences

Sentences with proscribed forms varied in botheration scores according to category with spelling inspiring the highest botheration rates, followed closely by morphology, syntax, and lexical items. Again, morphology and spelling have similar means and fall within the same Tukey HSD grouping (Figure 7).

Table 4: Mean Botheration Scores by Category for Proscribed Sentences

Category	*N*	*M*	*SD*
Lexical	35	3.09	1.560
Morphology	179	3.91	1.552
Spelling	170	3.95	1.497
Syntax	90	3.64	1.582
Total/Ave	**474**	**3.65**	**1.547**

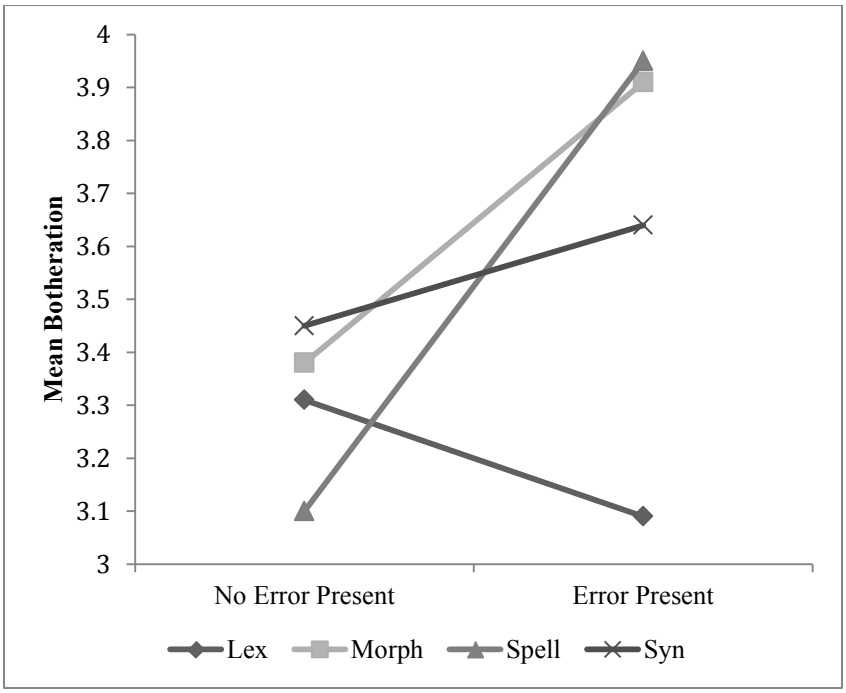
Figure 7: Mean Botheration Scores by Category for Proscribed Sentences with Tukey HSD Groups



The average mean botheration for proscribed sentences is higher than that for prescribed forms. An independent-samples T-test shows that this difference is significant, t (681) = 3.844, p = .000, calculating the mean for error present at 3.81 and the mean for no error at 3.32. Cohen's d is 0.29460, showing that this is a meaningful difference, though a small to moderate effect size (Cohen 1988). The differences between mean

botheration scores for prescribed and proscribed sentences in each category is illustrated in Figure 8.

Figure 8: Mean Botheration Scores for Prescribed and Proscribed Sentences by Category



This figure shows that the means are relatively close, and it illustrates that, surprisingly, the botheration mean for prescribed lexical items was higher than proscribed—suggesting that in some cases, participants were more bothered by a correct form than an incorrect form. (Please note that the scale in Figure 8 goes from 3 to 4, making the difference appear sharper than they actually are—recall that the botheration scale actually goes from 1 to 6. As discussed, the variability within botheration scores was high, and each mean fell between 3 and 4. The modified scale was used so that the trends and differences between the trends would be easier to see.)

The ANOVA showed that there was a significant difference in botheration scores for proscribed sentences by category, F (470, 1) = 8.610, R^2 = .023, p = 0.013. The Tukey HSD *post hoc* revealed that at the $p < .05$ level, as in the overall test, lexical items were significantly lower than morphological and spelling, but there was no significant difference between other groupings. It seems that combining the prescribed and proscribed forms together for a single analysis dilutes the effect of botheration by category, though the means are still similar—within the 3 to 4 range—and follow the same ranking pattern as combining all sentences. Table 4 and Figure 7 above show the mean botheration scores for proscribed sentences.

Item

Variation in botheration was next measured at the item level, which according to the mixed-effects model results, accounted for more of the variation than the categorical level.

All Sentences

Mean botheration scores for each item create a list ranking which items are most bothersome. Table 5 below shows this overall ranking of items, excluding the items with fewer than 10 responses (see Table 1 for these excluded items). Table 6 includes a ranking of most bothersome items within each category.

The ANOVA test showed a significant difference in item means for F (647, 1) = 3.452, R^2 = .157, p = .000. Note that the R^2 for item accounts for more of the variation in botheration scores than does the same score for category, though it is still fairly low, suggesting other latent variables. The Tukey *post hoc* test divided the items into four groupings, which are outlined in Table 5 and Table 6.

The top six most bothersome items and the lowest three are the only items in Tukey groupings that are significantly different from another group—in most cases, the top items are simply statistically more bothersome than the bottom ones, with a large spread of similar items in the middle.

Table 5:Ranking of Item Mean Botheration Score (All Sentences)

Rank	Category	Rule	*M*	*N*	*SD*	Tukey Group
1	Spelling	Except	4.74	23	1.054	A
2	Morphology	we was	4.68	31	1.249	AB
3	Syntax	multiple negation	4.63	24	1.279	AB
4	Morphology	Brang	4.49	35	1.541	ABC
5	Spelling	already	4.44	27	1.121	ABC
6	Morphology	Media	4.42	12	1.621	ABC
7	Syntax	Him and X as subject	4.19	26	1.550	ABCD
8	Spelling	all ready	4.09	32	1.553	ABCD
9	Spelling	emigrate	3.86	14	1.562	ABCD
10	Morphology	Data	3.82	11	1.722	ABCD
11	Morphology	Drunk	3.73	33	1.506	ABCD
12	Morphology	dived/dove	3.69	16	1.506	ABCD
13	Spelling	Effect	3.67	18	1.414	ABCD
14	Spelling	Affect	3.64	22	1.497	ABCD
15	Morphology	has went	3.56	18	1.653	ABCD
16	Morphology	Myself	3.55	11	1.368	ABCD
17	Lexical	nauseous	3.50	12	1.624	ABCD
18	Lexical	Literally	3.5	14	1.653	ABCD
19	Spelling	a lot/alot	3.47	17	1.700	ABCD
20	Syntax	either is	3.43	14	1.742	ABCD
21	Morphology	criteria	3.38	13	1.502	ABCD
22	Morphology	first/firstly	3.37	19	1.571	ABCD
23	Morphology	real/really	3.35	17	1.579	ABCD
24	Spelling	Alright	3.32	19	1.493	ABCD
25	Syntax	compound subject	3.31	13	1.109	ABCD
26	Spelling	a while	3.30	10	1.567	ABCD
27	Spelling	Allude	3.30	20	1.720	ABCD
28	Syntax	gerund with possessive	3.29	17	1.312	ABCD
29	Morphology	Slow	3.27	11	1.104	ABCD
30	Syntax	Between you and I	3.04	28	1.319	ABCD
31	Lexical	Farther	3	27	1.519	ABCD

32	Lexical	Further	2.91	11	1.221	ABCD
33	Morphology	bad as intensifier	2.83	24	1.204	ABCD
34	Morphology	these kind of	2.77	13	1.641	BCD
35	Syntax	Feel bad/badly	2.53	15	1.356	CD
37	Spelling	Awhile	2.44	16	1.365	D

Table 6: Ranking of Mean Botheration Scores of All Items by Category

Rank	Category	Rule	*M*	*N*	*SD*	Tukey Group
1	Lexical	nauseous	3.5	12	1.624	ABCD
2	Lexical	Literally	3.5	14	1.653	ABCD
3	Lexical	farther	3	27	1.519	ABCD
4	Lexical	Further	2.91	11	1.221	ABCD
1	Morphology	we was	4.68	31	1.249	AB
2	Morphology	Brang	4.49	35	1.541	ABC
3	Morphology	Media	4.42	12	1.621	ABC
4	Morphology	Data	3.82	11	1.722	ABCD
5	Morphology	drunk	3.73	33	1.506	ABCD
6	Morphology	dived/dove	3.69	16	1.506	ABCD
7	Morphology	has went	3.56	18	1.653	ABCD
8	Morphology	myself	3.55	11	1.368	ABCD
9	Morphology	criteria	3.38	13	1.502	ABCD
10	Morphology	first/firstly	3.37	19	1.571	ABCD
11	Morphology	real/really	3.35	17	1.579	ABCD
12	Morphology	Slow	3.27	11	1.104	ABCD
13	Morphology	bad as intensifier	2.83	24	1.204	ABCD
14	Morphology	these kind of	2.77	13	1.641	BCD
1	Spelling	except	4.74	23	1.054	A
2	Spelling	already	4.44	27	1.121	ABC
3	Spelling	all ready	4.09	32	1.553	ABCD
4	Spelling	emigrate	3.86	14	1.562	ABCD
5	Spelling	effect	3.67	18	1.414	ABCD
6	Spelling	affect	3.64	22	1.497	ABCD
7	Spelling	a lot/alot	3.47	17	1.7	ABCD
8	Spelling	alright	3.32	19	1.493	ABCD
9	Spelling	a while	3.3	10	1.567	ABCD
10	Spelling	allude	3.3	20	1.72	ABCD
11	Spelling	awhile	2.44	16	1.365	D
1	Syntax	multiple negation	4.63	24	1.279	AB
2	Syntax	Him and X as subject	4.19	26	1.55	ABCD
3	Syntax	either is	3.43	14	1.742	ABCD

4	Syntax	compound subject	3.31	13	1.109	ABCD
5	Syntax	gerund with possessive	3.29	17	1.312	ABCD
6	Syntax	Between you and I	3.04	28	1.319	ABCD
7	Syntax	Feel bad/badly	2.53	15	1.356	CD

These tables show that the spelling item *except* was, overall, the most bothersome item in this survey. It's interesting to note that both the highest and lowest items came from the spelling category. The top six most bothersome items contain dialectal issues, as was expected, though only three. *Except*, ranked first; *already*, another spelling item, ranked fifth; and *media*, a morphological item ranked sixth.

These top items do indeed show the pull of category—all fall within spelling or morphology, (with the exception of multiple negation in the syntax category) and may be the items that give these two categories their higher mean botherations. All lexical items fall within the middle range, though it is significantly the lowest group overall. The ranking by category in Table 6 shows clearly the varying number of items in each category that received a botheration score. There were only four lexical items that were even recognized as errors and given a botheration score by at least 10 people. Morphology has 14 items—all but one item received a botheration score—and spelling has 11, with syntax at 7. This seems to mirror the mean botheration by category rankings in the previous section: more spelling and morphology items were bothersome, with much fewer lexical items. This suggests that category is not a factor to be ignored, though the statistics suggest it is not a hugely important factor.

While spelling and morphology are higher categories in terms of botheration scores overall, it may well be that the categories are high because they happen to contain these particularly bothersome items (especially the dialectal markers)—and not

necessarily that the items rank high because they are within the categories of spelling and morphology. For example, the lowest item, *awhile*, falls in the spelling category—again suggesting the supremacy of individual item over category. (Interestingly, this item ranks much lower than its alternative spelling, *a while*—clearly a more marked, noticeable, and perhaps bothersome variation).

Prescribed Sentences

The following table includes a ranking of items by mean botheration for only prescribed sentences. Note that many of the Ns fall below 10—these were not removed for this analysis (though those items with fewer responses than 10 overall *are* still excluded, as in previous analyses). The ANOVA revealed no significant difference among these groups, with F (168, 1) = .1.262, p = .176.

Table 7: Ranking of Prescribed Items by Mean Botheration Score

Rank	Category	Rule	*M*	*N*	*SD*
1	Spelling	a while	4.67	3	.577
2	Spelling	already	4.50	2	.707
3	Morphology	Media	4.25	8	1.753
4	Lexical	nauseous	4.14	7	1.464
5	Spelling	except	4.00	1	0
6	Syntax	compound subject	4.00	5	1.000
7	Spelling	emigrate	3.86	7	1.676
8	Morphology	dived/dove	3.79	14	1.762
9	Morphology	Data	3.60	10	1.647
10	Morphology	criteria	3.60	5	.894
11	Morphology	has went	3.57	7	1.272
12	Morphology	Drunk	3.46	13	1.713
13	Lexical	farther	3.21	14	1.528
14	Syntax	gerund with possessive	3.18	11	1.328
15	Spelling	alright	3.08	13	1.498
16	Spelling	Affect	3.00	3	1.000
17	Morphology	myself	3.00	1	0
18	Syntax	either is	3.00	8	1.604
19	Morphology	first/firstly	3.00	5	1.581

20	Lexical	Further	3.00	1	0
21	Syntax	between you and I	2.92	13	1.115
22	Morphology	bad as intensifier	2.82	11	1.401
23	Spelling	Allude	2.80	5	1.304
24	Lexical	Literally	2.71	7	1.254
25	Morphology	Brang	2.67	6	1.366
26	Morphology	these kind of	2.67	6	1.862
27	Spelling	awhile	2.57	7	1.618
28	Morphology	real/really	2.50	2	.707
29	Spelling	all ready	2.17	6	1.602
30	Syntax	Him and X as subject	2.00	1	0
31	Spelling	Effect	1.00	1	0
32	Syntax	multiple negation	0.00	0	0.000
33	Morphology	we was	0	0	0
34	Spelling	a lot/alot	0	0	0
35	Morphology	slow	0	0	0
36	Syntax	Feel bad/badly	0	0	0

Proscribed Sentences

Because one of these items only has one botheration score (data), the *post hoc* test could not be performed. When this one item was removed, the ANOVA showed that there are significant differences between mean botheration scores among proscribed items, with $F(438, 1) = 3.632$, $R^2 = .220$, $p = .000$. Note that again, the R^2 emphasizes that item accounts for more variation than does category—and, specifically, proscribed items. While the ANOVA found significant differences, the Tukey HSD *post hoc* test did not detect any differences between groups, likely because of the large number of pairwise comparisons. However, an independent-samples *t* test comparing the item with the highest botheration (*brang*; m = 4.86) and lowest (*awhile*; m = 2.33), showed that there was a significant difference between these two, with $t(36) = 5.157$, $p = .000$.

The following tables show a ranking of proscribed items by mean botheration, first overall and then by category.

Table 8: Ranking of Proscribed Items by Mean Botheration Score

Rank	Category	Rule	*M*	*N*	*SD*
1	Morphology	data	6.00	1	0
2	Morphology	brang	4.86	29	1.302
3	Spelling	except	4.77	22	1.066
4	Morphology	media	4.75	4	1.500
5	Morphology	we was	4.68	31	1.249
6	Syntax	multiple negation	4.63	24	1.279
7	Spelling	all ready	4.54	26	1.174
8	Spelling	already	4.44	25	1.158
9	Lexical	Literally	4.29	7	1.704
10	Syntax	Him and X as subject	4.28	25	1.514
11	Syntax	either is/are	4.00	6	1.897
12	Morphology	drunk	3.90	20	1.373
13	Spelling	emigrate	3.86	7	1.574
14	Spelling	alright	3.83	6	1.472
15	Spelling	effect	3.82	17	1.286
16	Spelling	affect	3.74	19	1.558
17	Morphology	myself	3.60	10	1.430
18	Morphology	has went	3.55	11	1.916
19	Morphology	first/firstly	3.50	14	1.605
20	Syntax	gerund with possessive	3.50	6	1.378
21	Spelling	a lot/alot	3.47	17	1.700
22	Morphology	real/really	3.47	15	1.642
23	Spelling	allude	3.47	15	1.846
24	Morphology	slow	3.27	11	1.104
25	Morphology	criteria	3.25	8	1.832
26	Syntax	Between you and I	3.13	15	1.506
27	Morphology	dived/dove	3.00	2	2.828
28	Lexical	Further	2.90	10	1.287
29	Syntax	compound subject	2.88	8	.991
30	Morphology	these kind of	2.86	7	1.574
31	Morphology	bad as intensifier	2.85	13	1.068
32	Lexical	farther	2.77	13	1.536
33	Spelling	a while	2.71	7	1.496
34	Lexical	nauseous	2.60	5	1.517
35	Syntax	Feel bad/badly	2.53	15	1.356
36	Spelling	awhile	2.33	9	1.225

Table 9: Ranking of Proscribed Items According to Category

Rank	Category	Rule	Botheration	N	Std Dev
1	Lexical	Literally	4.29	7	1.704
2	Lexical	Further	2.90	10	1.287
3	Lexical	farther	2.77	13	1.536
4	Lexical	nauseous	2.60	5	1.517
1	Morphology	data	6.00	1	0
2	Morphology	brang	4.86	29	1.302
3	Morphology	media	4.75	4	1.500
4	Morphology	we was	4.68	31	1.249
5	Morphology	drunk	3.90	20	1.373
6	Morphology	myself	3.60	10	1.430
7	Morphology	has went	3.55	11	1.916
8	Morphology	first/firstly	3.50	14	1.605
9	Morphology	real/really	3.47	15	1.642
10	Morphology	slow	3.27	11	1.104
11	Morphology	criteria	3.25	8	1.832
12	Morphology	dived/dove	3.00	2	2.828
13	Morphology	these kind of	2.86	7	1.574
14	Morphology	bad as intensifier	2.85	13	1.068
1	Spelling	except	4.77	22	1.066
2	Spelling	all ready	4.54	26	1.174
3	Spelling	already	4.44	25	1.158
4	Spelling	emigrate	3.86	7	1.574
5	Spelling	alright	3.83	6	1.472
6	Spelling	effect	3.82	17	1.286
7	Spelling	affect	3.74	19	1.558
8	Spelling	a lot/alot	3.47	17	1.700
9	Spelling	allude	3.47	15	1.846
10	Spelling	a while	2.71	7	1.496
11	Spelling	awhile	2.33	9	1.225
1	Syntax	multiple negation	4.63	24	1.279
2	Syntax	Him and X as subject	4.28	25	1.514
3	Syntax	either is	4.00	6	1.897
4	Syntax	gerund with possessive	3.50	6	1.378
5	Syntax	Between you and I	3.13	15	1.506
6	Syntax	compound subject	2.88	8	.991
7	Syntax	Feel bad/badly	2.53	15	1.356

The Effect of Item and Sentence

A mixed-effects model was also performed to determine the importance of category, item, sentence, and prescribed/proscribed form to the overall model. Category and prescribed/proscribed (error present) were set as fixed effects, with sentence nested under item as random effects. The first model (Figure 9) included all of these effects and suggested that item accounted for more variation (SD = 0.4638) than sentence (SD= 0.2263) The residual was 1.4759, however, suggesting that latent variables account for much more of the variation than these effects. The simplest and final model (Figure 10) revealed that the effects that accounted for significant variation in botheration included error present (SD = 0.2738) and item (SD = 0.5133); R = 1.4902.

Figure 9: First Mixed-Effects Model

```
model_1 <- lmer(Botheration ~ Category + ErrorPres +
(1|Item/Sentence), data=botheration, REML = FALSE)
```

Figure 10: Final Mixed-Effects Model

```
model_5 <- lmer(Botheration ~ ErrorPres + (1|Item),
data=botheration, REML = FALSE)
```

The mixed-effects model results, like the results in the discussion above, show that while item and error present account for some variation, they only account for very little. There are other unseen variables at work—though this model shows that one of these was definitely not the sentence chosen to represent the test item. Participants would likely still be bothered by that item at the same rate, not matter which authentic sentences were used. However, the results also showed that category wasn't necessary to explain variation in botheration scores. These results provide some evidence for the importance

of individual preferences and “pet peeves” when it comes to the botheration of prescriptive rules.

Demographics

While not a major research concern, tests were also performed to determine the effect of the demographic gender, age, and education on botheration.

Gender

An independent-samples *t* test was performed to compare the botheration scores reported by men and those reported by women, regardless of prescribed and proscribed form. Mean botheration scores for men was 3.53 (SD = 1.569) and for women was 3.78 (SD = 1.536).The two means are statistically different, with $t(680) = 2.107$ and $p = .035$. These results suggest that women tend to report higher botheration scores than do men. However, Cohen’s *d* is only 0.16, which would be considered a small effect size (Cohen, 1988). This suggests that the significance may be due to the large sample size rather than the influence of the variable of gender, though this factor shouldn’t be ruled out entirely.

Age

A Pearson’s correlation was conducted to analyze the relationship between age group and reported botheration score. The age categories and composition of the participants can be reviewed in Figure 3 in the previous chapter. The correlation between age and botheration was not significant with $r = -.028$, $n = 683$, and $p = .465$. An ANOVA confirmed these results, with $F(676, 6) = 1.161$, $p = .325$. While it may be noteworthy that age wasn’t significant—it wouldn’t be too far off base to assume that the older generation complains more about the usage of the younger generation—the lack of

significance may in part be attributed to the large number of survey respondents in the 20 to 30 years old age bracket (42%).

Education

Education, however, appears to be a more significant variable influencing botheration scores, though perhaps not as strongly as might have been expected, considering that some claim prescriptive rules codify the use and preferences of the most educated speakers. (See Figure 4 in chapter 3 for a breakdown of education categories.) There was a positive correlation between botheration score and level of education, with $r = .153$, $n = 683$, and $p = .000$. However, this positive correlation, while significant, is very small and may again be the result of a large sample size.

An ANOVA suggests the importance of education level with $F(777, 6) = 4.982$, $R^2 = .037$, $p = .000$. The Tukey *post hoc* test revealed that the "some high school" group was significantly lower than the others, while the graduate degree and, interestingly, three years of college, levels were significantly higher. The following table shows mean botheration for each group with Tukey HSD groupings. (Note that the *N* may seem high—means were computed using every botheration score offered by every participant, so the same participant may be represented several times in the *N*. See Figure 4 for a breakdown of participant age groups.)

Table 10: Mean Botheration by Education Level

Education	*N*	Mean Botheration	Tukey HSD
Some high school	4	2.5	A
1 year of college	48	3.02	AB
High school	56	3.27	AB
2 years of college	151	3.61	AB
4+ years of college	265	3.63	AB
3 years of college	58	4.07	B
Graduate degree	101	4.16	B

Recognition

The results for error recognition will now be presented. Recall that for every sentence, participants were asked to identify whether or not an error is present in that sentence. The definition of a correct error identification was outlined in the Analysis of Results section of Chapter 3. The following section will follow an organization similar to the previous: first recognition will be discussed at the category level and then at the item level with all sentences, then prescribed and proscribed. Finally, the influence of demographics on recognition will be considered.

As discussed in Chapter 3, analyzing recognition with both prescribed and proscribed sentences together produces somewhat inflated results showing high rates of correct error recognition, as sentences with no error that were identified as having no error was considered a correct identification. Results at the category and item levels including all sentences are useful primarily for the purpose of comparison among the groups and less so in showing how often participants were able to recognize errors in the test sentences. Results for proscribed sentences, on the other hand, are more useful for this, and results for prescribed sentences identify areas where participants are objecting to correct forms.

Unlike botheration scores, a recognition score was given to every sentence on the survey, for a total of 3,822 scores (slightly higher than the anticipated 3,600—ten responses for each of 360 sentences. I'm not sure this happened, as MTurk was set to only allow ten people to take each survey. Only ten people were paid; these may have

shared the link with others). Because every item received at least ten scores, none were removed from the analysis.

Category

At the level of category, crosstabs and chi square tests were performed to explore error identification. Recognition seems to mirror botheration in terms of rank: For all sentences, spelling errors were most likely to be correctly identified, followed by morphology, syntax, and then lexicon (see Table 12 and Figure 11). Proscribed sentences follow the same pattern (see Table 13 and Figure 11).

Figure 11: Correct Error Identification by Category

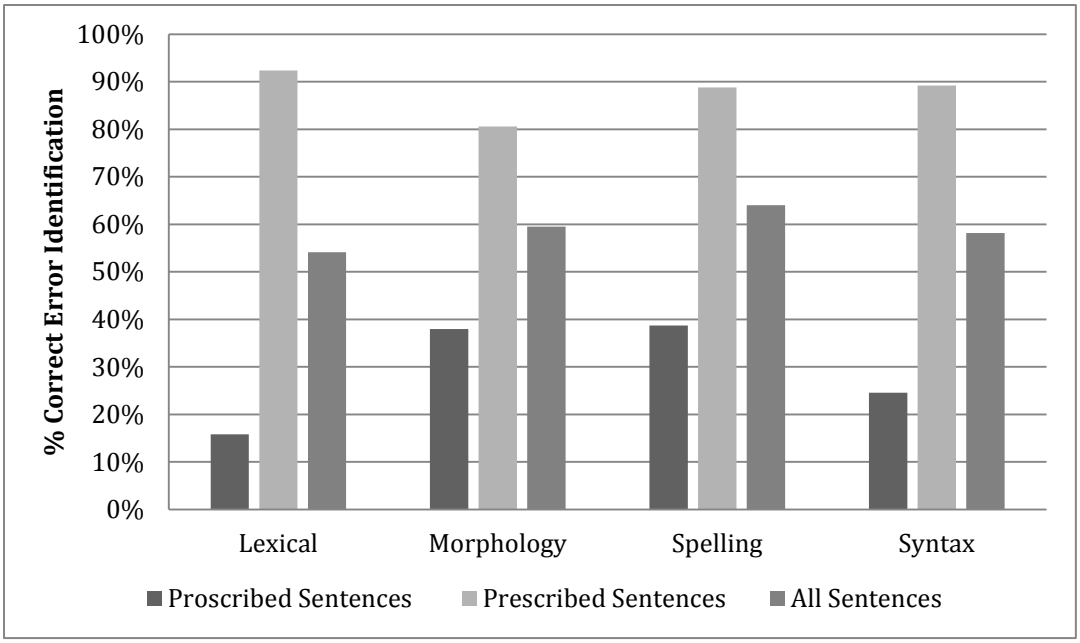


Table 11: Correct Error Identification by Category

Category	**All**	**Pro**	**Pre**
Lexical	15.80%	92.40%	54.11%
Morphology	38.00%	80.60%	59.50%
Spelling	38.70%	88.80%	64.00%
Syntax	24.60%	89.20%	58.20%

Prescribed sentences also have the highest recognition rate for spelling, followed by syntax, lexicon, and morphology. This means participants incorrectly identified correct forms as errors most often for the reverse order: morphology, lexicon, syntax, and spelling. It's interesting to note that spelling errors were most likely to be recognized as error, and sentences with correct forms of oft-misspelled words were more likely than other correct forms to be judged to have no error.

For all sentences, there is a statistically significant association between error recognition and category with $\chi(3) = 24.334$, $p = .000$.

Table 12: Crosstabs for Error Recognition by Category

		Category				Total
		Lexical	Morphology	Spelling	Syntax	
Incorrect ID	Count	436	389	350	396	1571
	Expected Count	389.7	394.2	398.7	388.4	1571.0
	% within Correct ID	27.8%	24.8%	22.3%	25.2%	100.0%
	% within Category	45.9%	40.5%	36.0%	41.8%	41.0%
	% of Total	11.4%	10.2%	9.1%	10.3%	41.0%
Correct ID	Count	514	572	622	551	2259
	Expected Count	560.3	566.8	573.3	558.6	2259.0
	% within Correct ID	22.8%	25.3%	27.5%	24.4%	100.0%
	% within Category	54.1%	59.5%	64.0%	58.2%	59.0%
	% of Total	13.4%	14.9%	16.2%	14.4%	59.0%
Total	Count	950	961	972	947	3830
	Expected Count	950.0	961.0	972.0	947.0	3830.0
	% within Correct ID	24.8%	25.1%	25.4%	24.7%	100.0%
	% within Category	100.0%	100.0%	100.0%	100.0%	100.0%
	% of Total	24.8%	25.1%	25.4%	24.7%	100.0%

For prescribed sentences, there is a statistically significant association between error recognition and category with $\chi(3) = 34.091$, $p = .000$. For proscribed sentences,

there is a statistically significant association between error recognition and category with $\chi(3) = 84.451$, $p = .000$.

Table 13: Error Recognition Crosstabs for Prescribed and Proscribed Sentences by Category

Error Present			Category				Total
			Lexical	Morphology	Spell	Syntax	
PRE	Incorrect ID	Count	36	94	55	53	238
		Expected Count	58.2	59.4	60.1	60.3	238.0
		% within Correct ID	15.1%	39.5%	23.1%	22.3%	100.0%
		% within Category	7.6%	19.4%	11.2%	10.8%	12.2%
		% of Total	1.9%	4.8%	2.8%	2.7%	12.2%
	Correct ID	Count	439	391	436	439	1705
		Expected Count	416.8	425.6	430.9	431.7	1705.0
		% within Correct ID	25.7%	22.9%	25.6%	25.7%	100.0%
		% within Category	92.4%	80.6%	88.8%	89.2%	87.8%
		% of Total	22.6%	20.1%	22.4%	22.6%	87.8%
	Total	Count	475	485	491	492	1943
		Expected Count	475.0	485.0	491.0	492.0	1943.0
		% within Correct ID	24.4%	25.0%	25.3%	25.3%	100.0%
		% within Category	100.0%	100.0%	100.0%	100.0%	100.0%
		% of Total	24.4%	25.0%	25.3%	25.3%	100.0%
PRO	Incorrect ID	Count	400	295	295	343	1333
		Expected Count	335.5	336.3	339.8	321.4	1333.0
		% within Correct ID	30.0%	22.1%	22.1%	25.7%	100.0%
		% within Category	84.2%	62.0%	61.3%	75.4%	70.6%
		% of Total	21.2%	15.6%	15.6%	18.2%	70.6%
	Correct ID	Count	75	181	186	112	554
		Expected Count	139.5	139.7	141.2	133.6	554.0
		% within Correct ID	13.5%	32.7%	33.6%	20.2%	100.0%
		% within Category	15.8%	38.0%	38.7%	24.6%	29.4%
		% of Total	4.0%	9.6%	9.9%	5.9%	29.4%
	Total	Count	475	476	481	455	1887
		Expected Count	475.0	476.0	481.0	455.0	1887.0

% within Correct ID	25.2%	25.2%	25.5%	24.1%	100.0%
% within Category	100.0%	100.0%	100.0%	100.0%	100.0%
% of Total	25.2%	25.2%	25.5%	24.1%	100.0%

Item

At the item level, frequency counts were performed to report rates of error recognition. The correct identification of errors will be ranked first by item within category (Table 15), and then by individual item (Table 15). These items are ranked according to correct identification of error in proscribed sentences. The tables are followed by charts that show percentage of correct item error recognition in each category.

For the sake of comparison, in Table 15 items are highlighted where the incorrect prescribed percentage is higher than the correct proscribed percentage. Cases where the two are close are highlighted as well. These are cases where the correct form was more often identified as an error than the incorrect form, or where both were identified as errors at similar rates. Such cases are interesting to note because they highlight instances where not only is a proscribed form becoming acceptable, but a prescribed form is itself inspiring language judgment, despite its “official” status as a “correct” variant. This concept will be discussed further in Chapter 5.

Table 14: Error Recognition Ranked by Correct Identification of Proscribed Sentences by Category

Category	Item	Proscribed		Prescribed		All	
		% Correct	% Incorrect	% Correct	% Incorrect	% Correct	% Incorrect
Lexical	farther	39.4	60.6	58.1	41.9	48.4	51.6
Lexical	Further	28.6	71.4	96.8	3.2	60.6	39.4
Lexical	infer	27.3	72.7	100	0	63.6	36.4
Lexical	Literally	22.6	77.4	75	25	49.2	50.8
Lexical	try and/try to	19.4	80.6	97	3	59.4	40.6
Lexical	nauseous	16.1	83.9	79.4	20.6	49.2	50.8
Lexical	Less	15.6	84.4	96.7	3.3	54.8	45.2
Lexical	adverse	6.5	93.5	100	0	52.5	47.5
Lexical	disinterested	6.5	93.5	100	0	54	46
Lexical	comprise	5.9	94.1	90	10	48	54.7
Lexical	among	3.3	96.7	93.8	6.3	50	50
Lexical	different than	3.2	96.8	100	0	53.1	46.9
Lexical	very unique	3.2	96.8	100	0	51.6	48.4
Lexical	between	0	100	100	0	50.0	50.5
Lexical	aggravate	0	100	100	0	50.8	49.2
Morphology	we was	100	0	100	0	100	0
Morphology	brang	93.5	6.5	81.8	18.2	87.5	12.5
Morphology	drunk	62.5	37.5	54.8	45.2	58.7	41.3
Morphology	real/really	46.9	53.1	93.5	6.5	69.8	30.2
Morphology	first/firstly	46.7	53.3	84.4	15.6	66.1	33.9
Morphology	bad as intensifier	41.9	58.1	73	27	58.8	41.2
Morphology	has went	35.3	64.7	84.4	15.6	59.1	40.9
Morphology	criteria	33.3	66.7	96.6	3.4	62.9	37.1
Morphology	slow	33.3	66.7	100	0	66.7	33.3
Morphology	myself	31.3	68.8	96.7	3.3	62.9	37.1
Morphology	these kind of	24.2	75.8	66.7	33.3	48	52
Morphology	media	15.6	84.4	71.9	28.1	34.9	56.3

Morphology	dived/dove	6.1	93.9	59.4	40.6	32.3	67.7
Morphology	hung/hanged	3.3	96.7	83.9	16.1	44.3	55.7
Morphology	data	3.1	66.7	67.7	32.3	34.9	65.1
Spelling	all ready	78.8	21.2	85.7	14.3	82.7	17.3
Spelling	already	71.9	28.1	93.5	6.5	82.5	17.5
Spelling	except	71.9	28.1	96.8	3.2	84.1	15.9
Spelling	affect	61.3	38.7	90.3	9.7	75.8	24.2
Spelling	a lot/alot	56.3	43.8	100	0	78.1	21.9
Spelling	effect	48.6	51.4	96.8	3.2	71.2	28.8
Spelling	allude	44.1	55.9	88.1	11.9	68.4	31.6
Spelling	awhile	28.1	71.9	77.4	22.6	52.4	47.6
Spelling	a while	23.3	76.7	90.6	9.4	58.1	41.9
Spelling	alright	23.3	76.7	61.3	38.7	42.6	57.4
Spelling	emigrate	21.9	78.1	77.4	22.6	49.2	50.8
Spelling	all together	15.2	84.8	87.1	12.9	50	50
Spelling	discrete	12.1	87.9	90.3	9.7	50	50
Spelling	discreet	10	90	96.8	3.2	54.1	45.9
Spelling	principle	9.4	90.6	100	0	55.4	44.6
Syntax	Him and X	83.3	16.7	97.1	2.9	90.8	9.2
Syntax	multiple negation	71.4	28.6	97.6	2.4	88.9	11.1
Syntax	Feel bad/badly	50	50	100	0	74.6	25.4
Syntax	Between you and I	38.7	61.3	64.5	35.5	51.6	48.4
Syntax	either is	27.3	72.7	57.5	42.5	46.8	53.2
Syntax	reason is because	24.2	75.8	100	0	60.9	39.1
Syntax	gerund with possessive	20	80	63.3	36.7	41.7	58.3
Syntax	compound subject	19.4	80.6	83.9	16.1	46.8	53.2
Syntax	stranded preposition	12.9	87.1	96.8	3.2	54.8	45.2
Syntax	singular “they”	9.7	90.3	94.1	5.9	53.8	46.2

Syntax	who/whom	9.4	90.6	90.3	9.7	49.2	50.8
Syntax	that/which	9.1	90.9	100	0	53.1	46.9
Syntax	who/that	9.1	90.9	100	0	52.4	47.6
Syntax	only	6.1	93.9	96.9	3.1	50.8	49.2
Syntax	split infinitive	0	100	100	0	50	50

Table 15: Error Recognition Ranked by Correct Identification of Proscribed Sentences

Category	Item	Proscribed		Prescribed		All	
		% Correct	% Incorrect	% Correct	% Incorrect	% Correct	% Incorrect
Morphology	we was	100.0	0.0	100.0	0	100.0	0.0
Morphology	brang	93.5	6.5	81.8	18.2	87.5	12.5
Syntax	Him and X as subject	83.3	16.7	97.1	2.9	90.8	9.2
Spelling	all ready	78.8	21.2	85.7	14.3	82.7	17.3
Spelling	already	71.9	28.1	93.5	6.5	82.5	17.5
Spelling	except	71.9	28.1	96.8	3.2	84.1	15.9
Syntax	multiple negation	71.4	28.6	97.6	2.4	88.9	11.1
Morphology	drunk	62.5	37.5	54.8	45.2	58.7	41.3
Spelling	affect	61.3	38.7	90.3	9.7	75.8	24.2
Spelling	a lot/alot	56.3	43.8	100.0	0	78.1	21.9
Syntax	Feel bad/badly	50.0	50.0	100.0	0	74.6	25.4
Spelling	effect	48.6	51.4	96.8	3.2	71.2	28.8
Morphology	real/really	46.9	53.1	93.5	6.5	69.8	30.2
Morphology	first/firstly	46.7	53.3	84.4	15.6	66.1	33.9
Spelling	allude	44.1	55.9	88.1	11.9	68.4	31.6
Morphology	bad as intensifier	41.9	58.1	73.0	27.0	58.8	41.2
Lexical	farther	39.4	60.6	58.1	41.9	48.4	51.6
Syntax	Between you and I	38.7	61.3	64.5	35.5	51.6	48.4
Morphology	has went	35.3	64.7	84.4	15.6	59.1	40.9

Morphology	criteria	33.3	66.7	96.6	3.4	62.9	37.1
Morphology	slow	33.3	66.7	100.0	0	66.7	33.3
Morphology	myself	31.3	68.8	96.7	3.3	62.9	37.1
Lexical	Further	28.6	71.4	96.8	3.2	60.6	39.4
Spelling	awhile	28.1	71.9	77.4	22.6	52.4	47.6
Lexical	infer	27.3	72.7	100.0	0	63.6	36.4
Syntax	either is	27.3	72.7	57.5	42.5	46.8	53.2
Morphology	these kind of	24.2	75.8	66.7	33.3	48.0	52.0
Syntax	reason is because	24.2	75.8	100.0	0	60.9	39.1
Spelling	a while	23.3	76.7	90.6	9.4	58.1	41.9
Spelling	alright	23.3	76.7	61.3	38.7	42.6	57.4
Lexical	Literally	22.6	77.4	75.0	25.0	49.2	50.8
Spelling	emigrate	21.9	78.1	77.4	22.6	49.2	50.8
Syntax	gerund with possessive	20.0	80.0	63.3	36.7	41.7	58.3
Lexical	try and/try to	19.4	80.6	97.0	3.0	59.4	40.6
Syntax	compound subject	19.4	80.6	83.9	16.1	46.8	53.2
Lexical	nauseous	16.1	83.9	79.4	20.6	49.2	50.8
Lexical	Less	15.6	84.4	96.7	3.3	54.8	45.2
Morphology	media	15.6	84.4	71.9	28.1	34.9	56.3
Spelling	all together	15.2	84.8	87.1	12.9	50.0	50.0
Syntax	stranded preposition	12.9	87.1	96.8	3.2	54.8	45.2
Spelling	discrete	12.1	87.9	90.3	9.7	50.0	50.0
Spelling	discreet	10.0	90.0	96.8	3.2	54.1	45.9
Syntax	singular "they"	9.7	90.3	94.1	5.9	53.8	46.2
Spelling	principle	9.4	90.6	100.0	0	55.4	44.6
Syntax	who/whom	9.4	90.6	90.3	9.7	49.2	50.8
Syntax	that/which	9.1	90.9	100.0	0	53.1	46.9
Syntax	who/that	9.1	90.9	100.0	0	52.4	47.6
Lexical	adverse	6.5	93.5	100.0	0	52.5	47.5

Lexical	disinterested	6.5	93.5	100.0	0	54.0	46.0
Morphology	dived/dove	6.1	93.9	59.4	40.6	32.3	67.7
Syntax	only	6.1	93.9	96.9	3.1	50.8	49.2
Lexical	comprise	5.9	94.1	90.0	10.0	48.0	54.7
Lexical	among	3.3	96.7	93.8	6.3	50.0	50.0
Morphology	hung/hanged	3.3	96.7	83.9	16.1	44.3	55.7
Lexical	different than	3.2	96.8	100.0	0	53.1	46.9
Lexical	very unique	3.2	96.8	100.0	0.0	51.6	48.4
Morphology	data	3.1	66.7	67.7	32.3	34.9	65.1
Lexical	aggravate	0	100.0	100.0	0	50.8	49.2
Lexical	between	0	100	100	0	50.0	50.0
Syntax	split infinitive	0	100.0	100.0	0	50.0	50.0

Figure 12: Correct Error Identifiction of Lexical Items

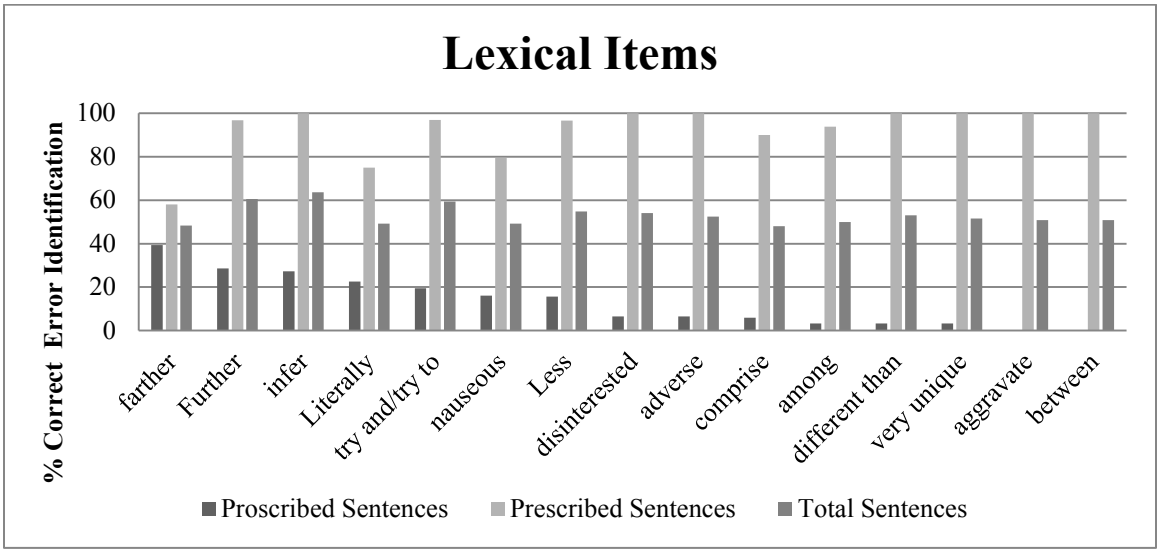


Figure 13: Correct Identification of Morphology Items

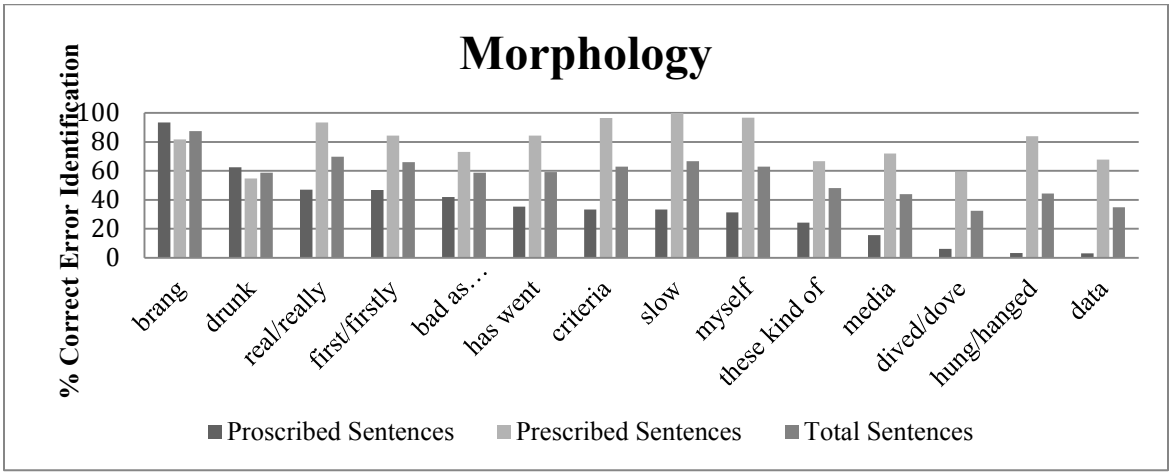


Figure 14: Correct Identification of Spelling Items

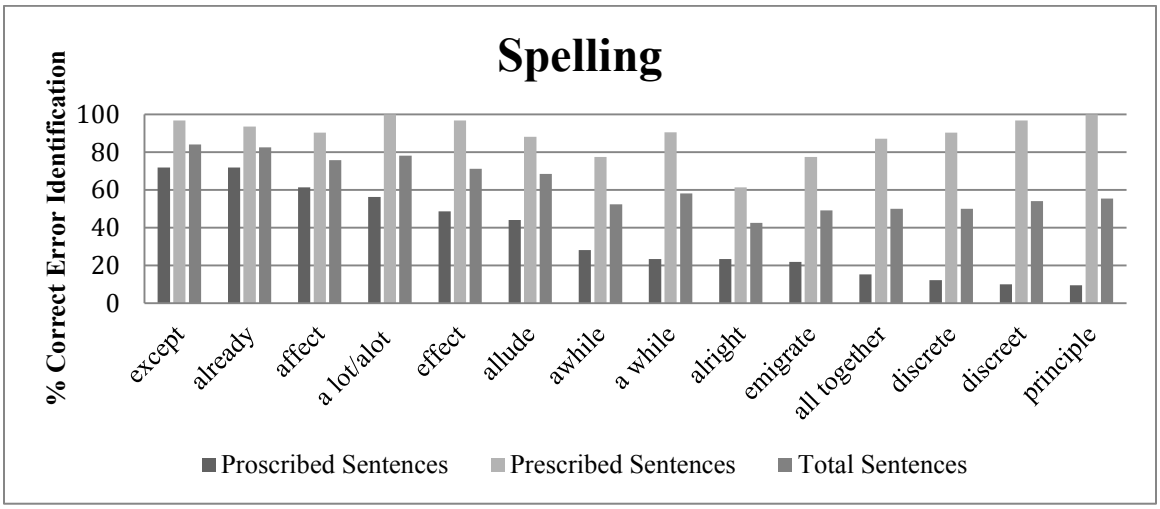
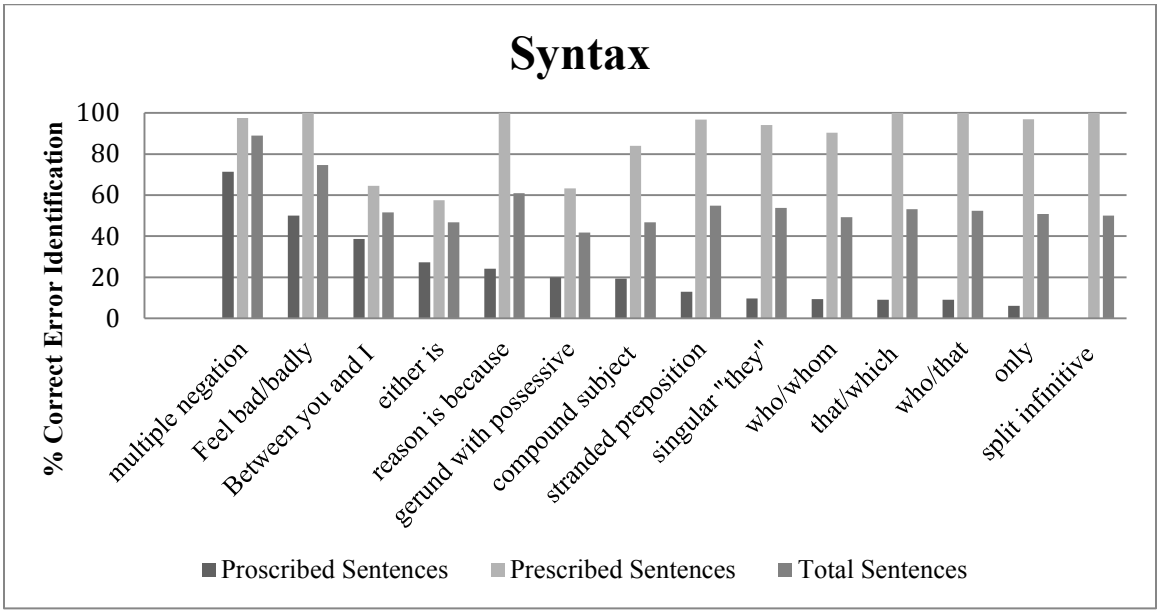


Figure 15: Correct Identification of Syntax Items



Demographics

Demographic data for correct error identification was analyzed with prescribed and proscribed sentences separated. Cross tabulations and chi square tests were performed for gender, age, and education.

Gender

For proscribed sentences, 287 men and 267 women correctly identified the error. A chi-square test showed no significant relationship, with $\chi(1) = 1.449$, $p = 0.229$. There was also no significant difference between the genders for prescribed. 767 women and 919 men correctly identified prescribed sentence as having no error—it appears that women were more likely to find an error in a correct sentence, but only at a level approaching significance, with $\chi(1) = 3.480$, $p = 0.062$.

Age

For proscribed sentences, there was a significant association between the demographic factor of age and correct error identification, for $\chi(6) = 25.822$, $p = 0.000$. The age group with the highest number of correctly identified errors was 20 to 30 (43.7%), followed by the 31 to 40 group (27.3%). These rankings most likely reflect the large number of participants in the 20 to 40 age group rather than a strong association due to age itself. For prescribed sentences, there was no association: $\chi(6) = 11.085$, $p = 0.086$, though age followed a similar ranking.

Education

Again, proscribed sentences contained a significant association between education and error recognition ($\chi[6] = 19.328$, $p = 0.004$) while prescribed did not ($\chi[6] = 5.999$, $p = 0.423$). For both groups, the 4+ years of college group had the highest percentage of correct identifications—39% for prescribed and 39.% percent for proscribed. Again, this could be because this education level had the highest number of participants overall.

Chapter 5: Discussion and Conclusion

This final chapter will include a discussion of the results, first addressing the primary research questions about the relationship between formal category and botheration and recognition of prescriptive rules. Next interesting cases of individual test items will be discussed. Improvements on these studies, as well as limitations and ideas for further research, will then be explored.

The results presented in the previous chapter provide evidence for several key ideas. First, as the prescriptivists would argue, the results showed that people notice and are bothered by violations of prescriptive rules. Participants noted and judged these violations, sometimes harshly. However, the results also provide evidence that some prescriptive rules are more meaningful than others, and some are barely, if at all noticeable. Additionally, for some rules, participants even favored the proscribed form over the "correct" prescribed form.

Research Questions: The Effect of Formal Category

The initial question guiding this thesis was whether or not some prescriptive rules, or some types of prescriptive rules, are more important than others. This study attempted to answer this question by finding out if some types of rules are more recognizable or bothersome than others—specifically, it set out to explore the relationship between the formal category a prescriptive rule falls under and how bothersome or recognizable a violation of that rule is.

Previous literature, as discussed in Chapter 2, provided some evidence that category is important. The botheration researchers often attempted to categorize their

results, and Chapman's (unpublished) meta study of the previous botheration results suggested that some types of errors are more bothersome than others, and these can be characterized by four formal categories, with morphological errors being the most bothersome, followed by spelling, syntax, and lexical errors. The current study set out to discover if this is the case—if some categories of errors are more significantly bothersome than others, and if they rank in this order. If that was the case, then perhaps generalizations could me made—*i.e.* any prescriptive rule could be assumed (to some extent) to be more or less bothersome based on which category it falls into.

Results, as outlined in Chapter 4, do show some variation by category in both botheration and recognition. Again, when all sentences were considered, spelling items were the most bothersome, closely followed by morphological items, then syntax, and finally lexical. However, the differences between the mean botheration scores, while significant, were not large. Lexical items were significantly lower than the other three categories, and spelling and morphology had nearly the same mean botheration score. There was no significant difference among the spelling, syntax, and morphology botheration scores. Recognition, or correct identification of error, mirrored botheration with violations of spelling rules the most often correctly identified, followed by morphology, syntax, and lexical. The *Chi*-square results showed that there was a significant difference between the expected and actual counts of correctly identified errors.

When it comes to botheration, the differences among the groups were not enough to support the hypothesis that the category into which a prescriptive rule falls could be used to generalize how bothersome or recognizable a violation of that rule might be. In

fact, the final mixed-effects model showed that category was unnecessary to help explain variation. And the effect size on the ANOVA suggested the effect of category was very small.

However, while generalization into the four groups doesn't seem possible given these results, category is not entirely to be ignored—especially in the case of lexical word meaning. The lexical category was the only group significantly different than the others, and only four items from this category were actually recognized as having an error by more than ten people (see Table 6)—and two of these are two forms of the same rule (*further* and *farther*). The recognition for proscribed lexical forms and functions was extremely low—only 13.5% of respondents correctly identified an error when there was one. The lexical items of *aggravate* and *between* used "incorrectly" weren't even recognized at all (see Table 1). It seems that prescriptive rules assigning meaning to a word tend to be less bothersome and less recognizable than other kinds of rules. While the quantitative results don't suggest a strong enough association for generalization, the evidence is there—generally, prescriptive rules assigning word meaning matter less to English speakers as their violations are less often recognized and least bothersome.

It seems that subtle differences in word meaning are opaque, if not meaningless. For example, in the categories of morphology, spelling, and syntax, the mean botheration scores for proscribed forms were higher than that for prescribed forms, while in the lexical category, the opposite is true: the mean botheration scores for all prescribed items were actually slightly *more* bothersome than that for proscribed. This could have several possible explanations. One explanation is that perhaps participants recalled that there is a prescriptive rule regarding a particular word—*further* for example—but couldn't quite

recall what the rule is, *i.e.* whether or not *further* is used to indicate physical or metaphorical distance, and decided to indicate that they saw an error. Or perhaps some proscribed forms are becoming more acceptable, providing some evidence of language change or semantic shifts (these both will be discussed later in this chapter).

Overall, it is clear that English speakers seem to disregard, ignore, or pass over prescriptive rules governing lexical meaning. This makes sense: following some morphological and syntactic rules may matter to processing the meaning of the sentence, but many lexical rules do not. Using the example above, *further* used in a physical sense does not obscure the meaning of the sentence it appears in. *Disinterested* and *uninterested* in form could both mean unbiased or uncaring, and in this rule, as in many others, the differences are subtle and the intended meaning is made clear in context. The results from this study provide quantitative evidence that readers don't notice, and likely don't care, about rules carefully separating out subtle differences in word meaning. Grammar sticklers may see this as evidence that these rules need to be more carefully taught—however, the recognition and botheration scores show that lexical proscribed (and prescribed) forms simply do not trip up readers.

As noted in the literature review, usage guides generally do not suggest any hierarchy of prescriptive rules, beyond the concession that some proscribed usages are acceptable in informal situations. This holds true for lexical rules as well; for example, one guide concedes that subtle differences in meaning are "bewildering, though they give English its depth and exactitude in expression" (Batko 2008:183). However, the results from this study suggest that these subtle differences are all but meaningless to the educated native English speakers participating in this study. Batko, who writes that

recognizing subtle word meaning differences "separates the men and women from the boys and girls when it comes to polished speaking," lists *aggravate vs annoy* among these important distinctions—though not a single participant in this study noticed, much less was bothered by, *aggravate* used to mean *annoy*. By the same token, no participants considered the prescribed meaning of *aggravate*—to make worse—a problem, either. It seems that these results provide some evidence that *aggravate* can carry both meanings without a problem; context makes the intended meaning clear without confusing or tripping up the readers. The same goes for other items tested in this study, such as *between* and *among*, *different than, disinterested, very unique, adverse, less*, and more.

The results from this study do not make a strong case for generalization by prescriptive rule category. However, with the statistical similarities among the other three formal categories, perhaps prescriptive rules dealing with syntax and morphology could be folded into one category—grammar. Then to some extent researchers and language professionals can generalize prescriptive rules within three categories: spelling-, grammar-, and semantic-type rules, with items falling in the third category tending to be least important to readers.

The "Pet Peeve" Effect

Rather than suggesting that certain *types* or categories of rules are more bothersome or recognizable, the data seems to be pointing toward a "pet peeve" effect—that is, for whatever reason, individuals are sticklers about a specific rule. They remember it, recognize it, and are highly bothered by it. As Joseph Williams (1981) noted, these "pet peeves" vary by individual. This can be seen in the data in instances where items had low *N*s but high botheration scores. For example, Table 1 from Chapter

3, which is reprinted below as Weaker evidence for the "Pet Peeve" effect can be found in Table 16 in items like *very unique.* Although *very unique* has a low botheration score (2.0), only one person noted is proscribed use—one person remembered the obscure rule that *unique* is absolute and cannot be qualified with an adjective like *very*. Even though this person was not bothered, he or she retains and notices this rule, while many others do not.

Table 16, shows the items that were removed from the data for having fewer than ten respondents. The item *different than* has only one respondent with a rating of 6. *Principle, adverse,* and *among* all have three respondents and average ratings of 4.0 to 4.67. Weaker evidence for the "Pet Peeve" effect can be found in Table 16 in items like *very unique.* Although *very unique* has a low botheration score (2.0), only one person noted is proscribed use—one person remembered the obscure rule that *unique* is absolute and cannot be qualified with an adjective like *very*. Even though this person was not bothered, he or she retains and notices this rule, while many others do not.

Table 16: Items with Fewer than 10 Botheration Scores

Item	**Category**	***N***	***M***
aggravate	Lexical	0	0
between	Lexical	0	0
split infinitive	Syntax	0	0
different than	Lexical	1	6.0
very unique	Lexical	1	2.0
disinterested	Lexical	2	3.0
adverse	Lexical	2	4.5
among	Lexical	3	4.0
principle	Spelling	3	4.67
that/which	Syntax	3	1.33
only	Syntax	3	2.67
who/that	Syntax	3	1.33
comprise	Lexical	4	2.5
discreet	Spelling	5	3.4

singular they	Syntax	5	2.4
stranded preposition	Syntax	5	2.8
Less	Lexical	6	2.17
hung/hanged	Morphology	6	3.38
discrete	Spelling	7	2.86
try and/try to	Lexical	8	3.0
who/whom	Syntax	8	3.38
reason is because	Syntax	8	3.0
infer	Lexical	9	3.56
all together	Spelling	9	2.44

And Table 8 in Chapter 4, showing the ranking of proscribed items by mean botheration, is topped by *data* with a score of 6—but given that rating by only one person, meaning one person of the thirty or so who saw *data* used with a singular verb was bothered, and this person as extremely bothered. This table shows other items with high botheration and low *N*, including *media*, *literally*, and *either is/are*. Evidence of this can also be found even in the prescribed table, with few people being bothered by *a while*, *already,* and *except.*

Additionally, the standard deviation scores for botheration are relatively high. Table 2 and Table 5 show that standard deviations hover between 1 and 2—fairly large variations, considering the scale only runs from 1 to 6. I believe these high standard deviations can be explained in part by the "pet peeve" effect. It seems that participants agreed that some items were more bothersome than others, but did not entirely agree on which broken rules were the greatest offenders. Individual participant may have been a greater source of variation than the items themselves.

The mixed-effects model results showed that item and prescribed/proscribed were important factors in explaining variation in botheration scores—however, the effect size in this and other tests suggested that the independent variables included in this study (*i.e.*

category, item, prescribed/proscribed, sentence, and demographics) actually accounted for only a small portion of the botheration scores. There are any number of unknown latent variables—how a person was feeling when they took the survey, speed of completing it, time of day, hunger, *etc*. Perhaps one of these is each individual's preferences—some people are just better at recognizing prescriptive rules in general, and some people are better than others at recognizing specific rules. Because not every participant responded to the same sentences, it's not possible to include participants in the mixed effects model, but I suspect it would be a significant category. Regardless of item type or category, some people just pick up on certain rules that others don't. For some reason or another, a rule learned in grade school sticks in a reader's mind, and that reader picks up on it throughout life. It seems that while some rules may be becoming forgotten, a few readers retain it as a "pet peeve" of sorts that they care about, even as its violations are committed and ignored by others around them.

Individual Prescriptive Rules

The mixed-effects model results, as discussed above, suggested that a rule's formal category is relatively unimportant. However, it did show that the item level is an important factor in botheration. Results for both recognition and botheration on the item level have some interesting implications. The following section will outline a few of the observations that can be gleaned from results of this study on the level of individual prescriptive rules. However, these discussions are by no means the most or only significant points to be made from this data, which may provide a goldmine with evidence for further studies. Please refer the tables included in Chapter 4 for rankings of items based on mean botheration and correct recognition.

Dialectal Items

As noted elsewhere in this study, several dialectal items were included among the test items, primarily in the category of morphology. It was predicted that these items, as nonstandard forms rather than variations of standard English, would inspire high botheration scores. Table 5 and Table 6 in the previous chapter, which rank botheration means at the item level, show that dialectal items indeed scored high. The top five most bothersome items three contain dialectal issues, though a surprise spelling item, *except*, took the top spot, with *already*, another spelling item, ranking fifth.

There was a concern that the inclusion of so many dialectal items within the morphological category (though there are some in syntax as well) might artificially give weight to that category, and maybe it did. However, it still was not decidedly the highest category—spelling means were roughly the same, and it was not significantly different from the syntax category. In fact, this item ranking shows that one dialectal item falls surprisingly near the middle of the ranking—*has went* at 15, even lower than *except* and *accept*, a common spelling mistake. Interestingly, the spelling *alright*, which is somewhat of an orthographic "dialectal" marker, ranks fairly low.

Prescribed vs. Proscribed

One of the interesting yet unexpected results from this study is prescriptive rules where prescribed forms—the presumed "correct" forms—bothered participants more or were identified as errors more often than the proscribed forms. Recall that the mixed-effects model also showed that whether the proscribed or prescribed usage was used was a significant source of variation in botheration scores.

Table 17 below shows items where mean botheration scores for the prescribed form were higher than for the proscribed from. *Emigrate* is included on this list because both the prescribed and proscribed forms had the same mean botheration score, and *data* is included because while the proscribed mean is 6.0, there is only one respondent—compared to the 10 who found an error in the prescribed form.

Table 17: Prescribed Items with Higher Botheration Scores than Proscribed

Category	Item	Pre *M*	Pre *N*	Pro *M*	Pro *N*
Spelling	a while	4.67	3	2.71	7
Spelling	already	4.5	2	4.44	25
Lexical	nauseous*	4.14	7	2.6	5
Spelling	emigrate*	3.86	7	3.86	7
Morphology	dived/dove*	3.79	14	3.0	2
Morphology	data	3.6	10	6.0	1
Morphology	criteria	3.6	5	3.25	8
Lexical	farther*	3.21	14	2.77	13
Lexical	further*	3.0	1	2.9	10

Table 18 shows cases where the "correct" use was more often recognized as an error than the "incorrect" use. This can be seen by comparing incorrectly identified prescribed items—prescribed forms determined to be incorrect—with correctly identified proscribed forms—proscribed forms determined to be incorrect. The table includes cases where prescribed forms incorrectly identified as errors have a higher percentage than proscribed forms correctly identified as errors. Cases where the two scores were relatively close are included as well. Note that many items, as would be expected, are included on both Table 17 and Table 18.. Where this is the case, the item is marked with an asterisk..

Table 18: Prescribed Items Recognized as Errors at Higher Rates than Proscribed Items

Category	Item	% Incorrect Prescribed	% Correct Proscribed	Difference
Syntax	singular "they"	5.9	9.7	-3.8
Syntax	compound subject	16.1	19.4	-3.3
Syntax	between you and I	35.5	38.7	-3.2
Syntax	who/whom	9.7	9.4	0.3
Spelling	emigrate*	22.6	21.9	0.7
Lexical	literally	25	22.6	2.4
Lexical	farther*	41.9	39.4	2.5
Lexical	among	6.3	3.3	3
Lexical	comprise	10	5.9	4.1
Lexical	nauseous*	20.6	16.1	4.5
Syntax	either is	42.5	27.3	15.2
Spelling	alright	38.7	23.3	15.4
Morphology	dived/dove*	40.6	6.1	34.5

The instances where prescribed items are identified as errors at all are very interesting—why are participants picking, out of an entire sentence, the prescribed realization of the very form I am testing, and identifying it as an error? A few explanations were discussed in a previous section—namely that perhaps participants are overcorrecting and recognizing that a word or structure has a rule associated with it, but aren't sure what that rule is; or it may be that the proscribed use is simply becoming more acceptable and/or common than the prescribed use.

A notable case is the word *nauseous*. The prescriptive rule dictates that this word used be used to mean "causing nausea," not "feeling nausea." However, results show that in a real-world context, more people considered this "correct use" to be an error, and those who considered the prescribed form an error were much more bothered by it (M = 4.14) than were those who considered the proscribed form an error (M = 2.6). While this rule is still "in the books," it seems that it is becoming meaningless—even incorrect—to relatively educated native English speakers. This makes sense—in context, if a person were to say they are nauseous, without further information it would be acceptable and

clear to assume that the person is feeling sick, not admitting to making others feel sick. The results for this and other words may provide some evidence of semantic shift—the prescribed meaning of *nauseous* is shifting among English speakers, and they don't seem to be too bothered or tripped up by it. Another example of this from the data shown in the tables above is *dove*—participants don't seem to mind it supplanting *dived* as the past tense of *dive.*

Another interesting example from this group is *further* and *farther*. Clearly participants seem to be aware that a rule assigning metaphorical and literal meaning to these words exists—however, they don't seem to know what this rule is. Both *further* and *farther* are more bothersome in their prescribed forms, and are more often recognized as errors when used correctly than incorrectly. These results suggest some confusion regarding the application of this rule, perhaps prompting overcorrection. I was also surprised to see *emigrate* on these lists—it may appear here because readers are more accustomed to finding the word *immigrate* and its forms in the media, and so note and perhaps find fault with *emigrate*, even used in its prescribed function.

Pop-Culture Targets

As mentioned, I included several prescriptive rules in the study that can be classified as "pop culture" items. That is, they are popular targets in jokes about the "grammar rules" English majors are sticklers about—and they are oft cited as "mistakes." These items are the split infinitive (Parales-Escudero, 2001; Close, 1987), stranded preposition (Yanez-Bouza, 2008), *who* vs. *whom* (Arts, 1994), and *literally* (Nerlich and Chamizo, 2003). The first three have been around for a while, and the last seems to be making a relatively recent strong appearance in popular culture. Despite the popularity

and familiarity of rules like split infinitives and stranded prepositions, even the authors of usage guides tend to denounce their importance as prescriptive rules.

Despite the stereotypes and supposedly strong feelings surrounding these items, when these rules are violated they don't seem to be noticeable or bothersome. The first three syntactic items, in fact, were cut out from the data analysis because so few people were bothered by them. They also had low recognition levels (see Table 14). The results from this study show that, despite the admonitions of sticklers past and present, these rules don't seem to matter very much; they aren't hindering meaning enough to bother readers, or even enough to be noticed by them.

Literally, on the other hand, seems to remain notorious: the proscribed function has a mean botheration of 4.29 (*N*= 7), with the prescribed function mean at 2.71 (*N*= 7). It's interesting that an equal number of people considered the prescribed function an error as the proscribed. This may be an example of overcorrection, with participants recognizing that some people have a problem with *literally* and so default to considering it an error. Those selecting the proscribed use as an error seem to have more confidence in the rule, evidenced by the high botheration score, while those picking the prescribed form as an error may be less confident about their choice.

Relative Pronouns

Another interesting note from Weaker evidence for the "Pet Peeve" effect can be found in Table 16 in items like *very unique.* Although *very unique* has a low botheration score (2.0), only one person noted is proscribed use—one person remembered the obscure rule that *unique* is absolute and cannot be qualified with an adjective like *very*.

Even though this person was not bothered, he or she retains and notices this rule, while many others do not.

Table 16, which shows items with fewer than 10 botheration scores assigned, is the presence of all relative pronouns tested in this survey: *that*/*which* ($N = 3$, $M = 1.33$), *who/that* ($N = 3$, $M = 1.33$), and *who/whom* ($N = 8$, 3.38). These three also appeared in the lowest third of syntactic items for correct error identification. *Who/whom* has the highest *N* and botheration score, which while still low isn't surprising, given this item's status as a "pop-culture target" as discussed in the above section. The appearance of this rule in TV shows, comics, and elsewhere may have clued participants into noticing the form. While important to the writers of grammar handbooks—each of these rules about relative pronouns appears in all but two of the usage guides consulted for this study (see Appendix A and Appendix B)—it seems that the often subtle distinctions between these pronouns are all but lost on native English speakers.

Referring to the restrictive/nonrestrictive distinction, Strunk and White's (in)famous tiny tome declares that "it would be a convenience to all if the two pronouns were used with precision. The careful writer, watchful for small conveniences, goes *which*-hunting, removes the defining *whiches*, and by so doing improves his work" (1979:59). While the authors may be off on the main point—that "all" would be more convenience with careful separation of the relative pronouns—language professionals, or "careful writers," seem to head off on "which hunts." In a study of what changes copyeditors target on academic manuscripts, Owen found that the most common edit made was changing *which* to *that* (2013:56). Prescriptive rules governing relative pronouns may be important to editors and usage-guide authors, but the present study

provides no evidence that these rules are important for readers. This point suggests another approach to answering this thesis's guiding question: some types of rules are more important than others—and some types of rules are more important to some groups of language users than others.

Rule Pairs

As noted in Chapter 3, several rules I selected to test included two forms. An example discussed above in this section is the rule governing *further* and *farther*, which for this study was separated into two separate rules: one stating the function of the form *further*, and another regarding *farther*. The results for this pair were discussed above. The rest of these pairs will now be discussed, with reference to Table 19 below. It seems that for each pair, there was not a large difference between the botheration *N*s and *Ms* and correct error identification. Perhaps one form could have been tested to represent the entire rule, as I elected to do with other rules governing two forms, such as *except* and *disinterested*. It seems that these rules are indeed conceptualized as a single rule, rather than separate rules dictating the function for each form. The exception to this may be found in *further* and *farther* which, while having almost the same botheration *M*, differ in *N*—the *further* form seems much less recognizable in its proscribed form than *farther*. The largest difference can be found in *between* and *among*, with *between* not garnering a single botheration score or correct error identification. It appears that the misuse of *among* is the more offensive form of this rule.

Table 19: Comparison of Rule Pairs

Item	**Botheration *N***	**Botheration *M***	**Proscribed Correct ID %**
further	11	2.91	*28.6*
farther	27	3	*39.4*
already	27	4.44	*71.9*

all ready	32	3.67	*78.8*
effect	18	3.67	*48.6*
affect	22	3.64	*61.3*
discreet	5	3.4	*10*
discrete	7	2.86	*12.1*
between	0	0	*0*
among	3	4	*3.3*

Demographics

Demographic data for age, gender, and education was collected for survey participants, and was thus analyzed to see if it could account for any variation in botheration or recognition. The results showed that the demographic data, especially for gender, was not a very good explanation for this variation, while education—unsurprisingly—was somewhat important. While Hairston (1981) posited that women were more bothered than men, my results showed that gender wasn't an important factor for error recognition or botheration. Age did not affect botheration, and education had a significant positive correlation, but not a large one. It seemed that having some college experience vs. not having any college made more of a difference in botheration than how much college. Age and education also were important to recognition scores.

Improvements upon Previous Studies

The results from previous botheration studies can be compared to the results from this study by converting the botheration means to *z* scores. Appendix D contains a table reporting the *z* scores for items in this study that also appeared in previous botheration studies with the *z* scores calculated by Chapman (unpublished) for these same items. The study design and analysis of the present study draws upon the basic design of past botheration studies, though it differs and, ideally, improves upon them. The purpose of

this section is to review the most significant and deliberate variations and improvements. It should be noted that some of these differences are present, at least to some extent, in one or two of the previous studies.

Test Items and Sentences

The botheration researchers used constructed sentences on their surveys, with the exception of Kantz and Yates (1994), who drew test sentences from student papers. The present study uses authentic, published sentences with minimal editing. This adjusts for potential sentence awkwardness—respondents in previous studies may have been responding to unnatural wordings or usage—and provides an example of how proscribed forms might appear in a real-world context, the way participants might actually encounter them. Data showed that this approach worked: The results from a mixed-effects model, as described in the previous chapter, showed that individual sentence was an unimportant factor in variation in botheration scores. The actual errors were driving responses, not the sentences themselves, as might have happened if constructed or unnatural sentences interfered with clarity.

The prescriptive rules tested were selected from usage guides and previous research, while botheration-study authors seemed to use the rules they taught in class, rules they believed college students struggled with the most (*e.g.* Beason), and rules tested on previous surveys. The present study also selected items according to four formal categories and structured the surveys so that each category was evenly represented on each.

Error Identification

The present study offers participants the chance to identify which error they were responding to when they rate how much they are bothered by the error. This allows the study to isolate the test item in question and make sure data analyzed represents a response to this test item and not to sentence quirks or other issues. In preparing the data for analysis, I went through every response to check if the participant was actually responding to the test item I wanted to target. In the raw data, 34.6% of 3,830 responses identified sentences as containing errors, and after the data was corrected, only 21.3% of responses identified errors—about 500 original responses were not responding to test items and could have potentially altered my results. This is definitely something that could have happened to the botheration researchers who did not ask participants to identify what they were responding too, and thus they really have no idea how accurately their results measure responses to the rules they intended to study.

Participants

Participants were recruited from a general American audience, not from a convenience sample or from a sample of especially educated or powerful people (*e.g.* professors and employers). This study thus gauges the attitudes of "normal," everyday (though slightly better educated, it turned out) readers and consumers of language. While not vested in prescriptivism, these participants are nonetheless shaped and influenced by it.

Prescribed and Proscribed Forms

Unlike in previous studies, as many sentences contained no error as those that did. Of the 360 sentences that appeared over the 30 surveys, exactly half contained prescribed forms of the test items. Few botheration researchers contained both prescribed and proscribed

forms, and those who did (*e.g.* Kantz and Yates) included only limited examples, with the test dominated with proscribed forms.

Including so many prescribed forms, and warning participants that not every sentence will contain an error, slightly shifts the mind-set of the participant. In a way, it mitigates the test-taking, "error-hunting" environment. Instead of hunting through each sentence to find something that might be an error, they are evaluating whether or not a sentence actually contains an error. They do not feel forced to find an error in every sentence, and thus feel more comfortable when a sentence appears correct. This encourages them to avoid second-guessing their judgments and looking more closely to find an error that they normally wouldn't notice or be bothered by in a more natural situation. Also, including equal numbers of prescribed and proscribed forms allowed this thesis to evaluate which prescribed forms were bothersome, and which were more surprisingly more bothersome than the proscribed forms.

As mentioned above, in the present study only 34.6% of responses identified an error, with 21.3% responding to a test item—one might predict these numbers might be closer to 50%, given that half of the sentences on the survey contained errors. This suggests that readers struggle to recognize violations of prescriptive rules in natural sentences, even in a testing environment.

Each item in the present study was also tested multiple times—each appeared six times, three in prescribed form and three in proscribed. This helped isolate the variable of test item in the analysis, and helped the study evaluate reactions to varying forms of each rule.

Limitations

Even with these strengths, the present study has many limitations, though perhaps ones that can be simply corrected in further research. The demographics of the participants were a bit wanting in diversity. While relatively equal numbers of men and women took the survey, the data showed a large majority of participants in their 20s and 30s and with four-year college degrees. Perhaps this is simply the demographic that frequents Mechanical Turk; either way, it would have been helpful to have a more evenly distributed participant pool. Additionally, the Likert scale used runs from 1 to 6. I don't know if it was clear enough that 1 represented a recognition of the error but no botheration from it. This option—one of seeing an error but not feeling bothered by it—could have been a more explicit one by using a 0 instead of a 1 as the lowest item on the scale.

Other weakness are not so easy to fix. The artificial test setting, including a survey with official BYU linguistics department headings and instructions to find an error most definitely primed participants to read closely, hunting for mistakes. As is inherent in this environment, they read single sentences more closely than they would have had the sentences appeared in a natural context, where they would have been less likely to see and be bothered by error. However, this effect may have been mitigated somewhat by the desire of Mechanical Turk workers to finish tasks quickly. The faster they complete tasks, the more money they can make in an hour. Some participants completed the survey in two minutes or less—at this pace, they were likely quickly scanning sentences as they would in the real world. However, they could also have noticed that they have fewer questions to answer when they indicate that a sentence does not contain an error and might have been more inclined to select this option. The effect of the test environment

may also have been mitigated somewhat by the inclusion of prescribed forms, which, as discussed above, may have made participants feel more confident in declaring that a sentence contained no error.

Another potential limitation is the inclusion of dialectal items within the category of morphology. While this category is perhaps a good fit for these items, they definitely skewed botheration scores for the morphology category higher. In a sense, these dialectal items are a different kind of "violation" than those prohibited by prescriptive rules. Granted, sticklers would cringe to hear a double negative or a *we was*, but prescriptive rules tend to stay within the realm of guarding word meaning and spelling distinctions, variations that exist within, not without, standard English. Because this was not a study of nonstandard English, perhaps it would have been better to not include these items, or to put them in their own category, or to look for other similar nonstandard items to place in each of the other categories.

Additionally, the item categorization used in this study may have some flaws, with some items perhaps fitting in two categories—*i.e further* and *farther* were used as lexical items, but they might have been spelling items; spelling items that included spacing issues might have been considered syntax items. Also, it is likely that different categorical groupings beyond, or more subtle than, these four used in my study are possible and could better explain variation in mean botheration scores. A closer and more fine-tuned look at categorization could be a subject of a future study.

Future Studies

A wealth of future studies based on this data is open to interested researchers. There are many more ways the data in this study could be analyzed that were beyond the

scope of this thesis. The results of this study could be compared to those of previous botheration studies by using *z* scores (see Appendix D). Such an analysis could combine data into a larger meta-study of botheration by category or item, or it could show change in botheration for proscribed forms over time, as some of the botheration studies are thirty or more years old. A comparison could also show differences in botheration levels among the groups of professionals, academics, and a more general readership, represented by the Mechanical Turk workers.

Additionally, a corpus study would likely add greater insight into the results of this study, especially when it comes to pairs like *further* and *farther*. A corpus study might explain why some forms are more recognizable than others. Some forms with lower botheration scores might simply be more uncommon and thus more unfamiliar to readers. And for words like *nauseous*, a corpus search comparing *nauseous* and *nauseated* may show that the former is being used more frequently over time, thus supporting and providing more evidence for shifting acceptability of the proscribed form.

It may also be interesting to compare botheration and recognition results to how often the prescriptive rules appear in usage manuals, as collected by Chapman. The number of times each item used in this study appear in usage guides is reported in Appendix B, though the results have not been compared to these tallies in this thesis. Additionally, an appropriate continuation in this study of responses to error would be to examine what participants reported when they incorrectly identified "errors." Because participants identified what part of the sentence they believed to be in error, I was able to throw out responses that did not respond to a test item in this study. However, looking at

what types of structures bothered participants and how much, beyond intentionally planted "errors," might be profitable.

There is also potential for new studies to build upon the present research. Additional prescriptive items could be added beyond those tested here, even more obscure rules, to see if categorical rankings still hold, or to provide a more comprehensive ranking of botheration of individual rules. This study could be replicated with different sentences—which, according to the analysis, shouldn't change the results very much. Dialectal items could be removed, or more could be added and analyzed as a separate category.

The present study focused on written English—the sentences were taken from written registers and presented to participants in written form. The categorization of rules, and the rules themselves, were selected with a written, published register in mind. A similar study with spoken English would be an interesting continuation of research on judgments on variations in standard English. Many usages guides claim to help readers avoid censure in speaking, in addition to writing. A botheration study of spoken English might ask participants to listen to sentences containing a proscribed form and identify an error.

This study tested some prescriptive rules that governed more than one form or function, such as *further* and *farther*. In this example, both forms were tested in this study, but many forms, such as *except*, the most bothersome item, *disinterested*, and *principle*. A possible future study would zero in on rules such as these—rules governing or distinguishing two forms—and compare recognition, botheration, and even frequency with a complementary corpus study between the two forms.

Conclusion

With the conventions of prescriptivism considered "quite possibly the last repository of unquestioned authority for educated people in secular society" (Cameron 1995:29), the desire to avoid harsh judgment or censure for using proscribed forms is strong among English speakers. For this reason, prescriptive writers publish usage guides to "provide them with rules, remind them of certain conventions" to help writers "write so that what they have to say will be understood, respected, even enjoyed" (Ebbitt and Ebbitt 1990:v). And it seems that this help might be appreciated, as "linguistic bigotry is among the last publically expressible prejudices left . . . . intellectuals who would find it unthinkable to sneer at a beggar or someone in a wheelchair will sneer without compunction at linguistic 'solecisms'" (Cameron 1995:29). But are all linguistic offenses equally bothersome?

This thesis began with the question of whether or not the native speakers of English regard some types of these prescriptive rules as more "important" or "meaningful" than another. Given the abundance of prescriptive rules, it's not likely that all speakers would care—or even recognize—every one. The results of this study show quantifiably that some types of rules are more "important," in that they bother readers more and are more recognizable than other types of errors. And some are less important—particularly, rules separating semantic meanings. Table 20 below summarizes which items it seems people do care about—those with high recognition and botheration scores—and items that they do not recognize and are not bothered by. It may be possible to generalize by assuming that lexical prescriptive rules will be less important to a general educated American audience than spelling or grammar rules, and that

nonstandard dialectal forms will be even more bothersome. However, the ability to generalize these results is limited: there is some evidence for a "pet-peeve" effect. Individuals seem to simply be bothered by different rules, without strong patterns showing some types of rules sharply more important than others.

Table 20: Results Summary Chart

Items with High Botheration/Recognition:	
Spelling	except
Morphology	we was
Syntax	multiple negation
Spelling	already
Morphology	brang
Morphology	media
Syntax	him and X as subject
Spelling	all ready
Morphology	drunk
Spelling	affect
Spelling	alot
Items with Low Botheration/Recognition:	
Lexical	aggravate
Lexical	between
Lexical	split infinitive
Lexical	different than
Lexical	very unique
Lexical	disinterested
Lexical	adverse
Lexical	among
Spelling	principle
Spelling	that/which
Syntax	only
Syntax	who/that
Lexical	comprise
Spelling	discreet
Syntax	singular *they*
Syntax	stranded preposition
Lexical	less
Morphology	hanged/hung
Spelling	discrete
Lexical	try and/try to
Syntax	who/whom
Syntax	reason is because
Lexical	infer
Spelling	all together
Morphology	these kind of

Syntax	feel bad/badly
Spelling	awhile
Items With More Bothersome/Recognizable Prescribed forms	
Spelling	a while
Spelling	already
Lexical	nauseous
Spelling	emigrate
Morphology	dived/dove
Morphology	data
Morphology	criteria
Lexical	farther
Lexical	further
Syntax	singular *they*
Syntax	compound subject
Syntax	between you and I
Syntax	who/whom
Lexical	literally
Lexical	among
Lexical	comprise
Syntax	either is

This study shows that prescriptivists have a point—people do judge proscribed forms. They have their pet peeves, they pick up on violated rules and are bothered by them, sometimes severely. However, sometimes prescribed forms—"correct" forms—are judged as well. It seems that people will notice and be bothered by even simple variations in standard English; perhaps this is inevitable in a "culture of correctness." While the botheration of prescribed forms and low recognition rates for other violated rules may show some evidence of language change and the arbitrary nature of some rules, it's likely that some prescriptivists would find this evidence of further language decay. To them, "the fact that educated speakers and writers use objective *who* is merely evidence of the widespread nature of the error" (Battistella 2005:54).

While participants in this study were bothered and did recognize error, it seemed to happen at a relatively low rate. When exactly half of the sentences contained

proscribed forms, only about a third were recognized as having errors, with fewer being the errors included for the purposes of this study, and even less being proscribed forms. Usage guides may posit that they contain rules to help people avoid judgment, the truth may be the average reader does not know the rules or isn't bothered by their violations—their violations don't trip the readers up enough to draw attention—and that people simply have their own unpredictable pet peeves.

The results of this study may be useful to people who care about "correctness," or to an audience similar to those who wrote, published, and read the original botheration studies; namely, teachers, editors, job applicants, etc. To these people, these results offer a guiding generalization and a ranking of the most and least offensive terms, and evidence for the pet peeve effect. The study also shows that "correctness" is more complicated than a simple binary—responses to proscribed forms can be quantified on a scale.

This study also demonstrates how botheration studies can be used in a linguistic usage study. Botheration studies can be used to study and quantify usage, providing additional evidence and support to corpus studies, but also providing something that corpus data can hint at but not provide—quantifiable evidence of language user opinion and judgment.

References

Aarts, Flor. 1994. Relative who and whom: Prescriptive rules and linguistic reality. *American Speech* 69(1). 71–79.

Algeo, John. 1991. Sweet are the usages of diversity. *Word* 42. 1–17.

Andersson, Lars and Peter Trudgill. 1992. *Bad Language*. New York: Penguin Books.

Baron, Naomi S. and Rich Ling. 2011. Necessary smileys and useless periods. *Visible Language* 45(1/2). 45– 67.

Batko, Ann. 2008. *When bad grammar happens to good people: How to avoid common errors in English.* New York: Career Press.

Battistella, Edwin L. 2005. *Bad language: Are some words better than others?* New York: Oxford University Press.

Beal, Jean. 2010. The grocer's apostrophe: popular prescriptivism in the 21st century. *English Today* 26. 57–64.

Beal, Jean, Carmela Nocera, and Massimo Sturiale. 2008. *Perspectives on prescriptivism: Vol. 73. Linguistic insights: studies in language and communication.* Bern, NY: Peter Lang.

Beason, Larry. 2001. Ethos and error: How business people react to errors. *College Composition and Communication* 53. 33–64.

Burchfield, Robert W. 2004. *Fowler's Modern English Usage*. New York: Oxford University Press.

Cameron, Deborah. 1995. *Verbal hygiene: The politics of language*. London: Routledge.

Chomsky, Noam. 1986. *Knowledge of language: Its nature, origin, and use.* New York: Praeger.

Close, R. A. 1987. Notes on the split infinitive. *Journal of English Linguistics* 20(2). 217-229.

Cohen, Jacob. 1988. *Statistical power analysis for the behavioral sciences.* New York: Routledge.

Crystal, David. 2006. *The fight for English: How language pundits ate, shot, and left*. New York: Oxford University Press.

Drake, Glendon. F. 1977. *The role of prescriptivism in American linguistics 1820–1970*. Studies in the history of linguistics, 13. Amsterdam: John Benjamins.

Ebbitt, Wilma R. and David R. Ebbitt. 1990. *Index to English*. New York: Oxford University Press.

Fishman, Joshua A. 1982. *Sociolinguistics: A Brief Introduction*. New York: Newbury House.

Gilsdorf, Jeanette W. and Don J. Leonard. 2001. Big stuff, little stuff: A decennial measurement of executives' and academics' reactions to questionable usage elements. *The Journal of Business Communication* 38. 439–475.

Gray, Loretta. S. and Paula Heuser. 2003. Nonacademic professional's perception of usage errors. *Journal of Basic Writing* 22. 50–71.

Hairston, Maxine. 1981. Not all errors are created equal: Nonacademic readers in the professions respond to lapses in usage. *College English, 43,* 794–806.

Heath, Shirley Brice 1980. Standard English: Biography of a symbol. In Timothy Shopen and Joseph M. Williams (Eds.), *Standards and dialects in English*. 3–32. Cambridge, MA: Winthrop.

Johnson, David and Lewis VanBrackle. 2012. Linguistic discrimination in writing assessment: How raters react to African American "errors," ESL errors, and standard English errors on a state-mandated writing exam. *Accessing Writing* 17. 35–54.

Joseph, John Earl. 1987. Eloquence and power: The rise of language standards and *standard languages.* New York: Blackwell.

Kantz, Peggy, and Bob Yates. 1994. "Whose Judgments? A Survey of Faculty Responses to Common and Highly Irritating Writing Errors." Fifth Annual Conference of the NCTE Assembly for the Teaching of English Grammar. Normal, IL.

Leonard, Don J. and Jeanette W. Gilsdorf. (1990). Language and change: Academics' and executives' perceptions of usage errors. *The Journal of Business Communication* 27. 137–158.

Lowth, Robert. 1967. *A Short Introduction to English Grammar*. Ed. R. C. Alston. No. 18. Menston, England: The Scolar Press Limited.

Lynch, Jack. 2009. *The lexicographer's dilemma: The evolution of "proper" English, from Shakespeare to South Park.* New York: Walker & Co.

Manning, Alan D. 2002. The grammar instinct interface. *IEEE Transactions on Professional Communication* 45. 133–137.

Mason, Winter and Siddarth Suri. 2011. Conducting behavioral research on Amazon's Mechanical Turk. *Behavioral Research Methods.* 44. 1–23.

Merriam-Webster's dictionary of English usage. 1994. Place: Merriam-Webster.

Merrier, Patricia A. (1991). The extent to which students' reactions to usage errors change as a result of having completed a business communication course. *Business Communication Quarterly* 54. 27–30.

Miege, Guy. 1969. *The English Grammar*. Ed. R. C. Alston. No. 152. Menston, England: The Scolar Press Limited. 1969.

Milroy, James and Lesley Milroy. 1999. *Authority in language: Investigating standard English*. New York: Routledge.

Murray, Lindley. 1969. *English Grammar*. Ed. R. C. Alston. No. 106. Menston, England: The Scolar Press Limited.

Nerlich, Brigitte and Pedro J. Chamizo Dominguez. 2003. The use of literally. *Annual Review of Cognitive Linguistics* 1. 193–206.

Owen, Jonathan. 2013. *At the coal-face of standardization: Uncovering the role of copy editors in standardizing the English language.* Provo, Utah: Brigham Young University thesis.

Perales-Escudero, Moises. 2011. To split or to not split: The split infinitive past and present. *Journal of English Linguistics* 39(4). 313-334.

Percy, Carol. 2004. Consumers of correctness Men, women, and language in eighteenth-century classified advertisements. In *New Perspectives on English: Selected Papers from 12 ICEHL, Glasgow, 21-26 August 2002.* eds Smith, J. J., Horobin, S, and Kay, C. 2004. John Benjamins Publishing Company: Phildelphia.

Pinker, Steven. 1994. *The language instinct: How the mind creates language.* New York: William Morrow.

Shaw, Harry. 1986. *Errors in English and ways to correct them.* New York: Collins Reference.

Sprouse, Jon. 2010. A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavioral Research Methods* 43. 155–167.

Truss, Lynne. 2001. *Eats, Shoots & Leaves.* New York: Gotham Books.

Walsh, Bill. 2013. *Yes, I could Care Less: How to Be a Language Snob without Being a Jerk*. New York: St. Martin's Press.

Walsh, Bill. 2004. *The Elephants of Style*. New York: McGraw-Hill Education.

Wardhaugh, Ronald. 1999. *Proper English: myths and misunderstandings about language.* Malden, MA: Blackwell Publishers.

Williams, Joseph H. 1981 The phenomenology of error. *College Composition and Communication* 32. 152–168.

Yanez-Bouza, Nuria. 2008. To end or not to end a sentence with a preposition: An 18th-century debate. *Linguistic Insights - Studies in Language and Communication* 73.

Young, Thomas E. 1991. The subjunctive and you: What college professors think. *The English Journal, 80,* 38–41.

Appendixes

Appendix A

Usage Guides Consulted in Selecting Test Items

Author/Editor	**Title**	**Year**
Copperund, Roy H.	American Usage: The Consensus	1970
Merriam-Webster, Inc.	Merriam-Webster's Dictionary of English Usage	1989
Ebbitt, Wilma R. and David R. Ebbitt	Index to English	1990
Harry Shaw	Errors in English and Ways to Correct Them	1993
Kenneth G. Wilson	The Columbia Guide to Standard American English	1993
American Heritage Dictionaries	The American Heritage Book of English Usage	1996
R. W. Burchfield	Fowler's Modern English Usage	1996
	Common Errors	1997
Bill Walsh	Lapsing Into a Comma: A Curmudgeon's Guide to the Many Things that Can Go Wrong in Print--and How to Avoid Them	2000
Thomas Parrish	The Grouchy Grammarian: A How-Not-To Guide to the 47 Most Common Mistakes in English Made by Journalists, Broadcasters, and Others Who Should Know Better	2002
Patricia T. O'Conner	Woe is I: The Grammarphobe's Guide to Better English in Plain English	2003
Bryan A. Garner	Gardner's Modern American Usage	2003
James R. Cochrane	Between You and I: A Little Book of Bad English	2004
Ann Batko	When Bad Grammar Happens to Good People: How to Avoid Common Errors in English	2004
Mark Lester and Larry Beason	The McGraw-Hill Handbook of English Grammar and Usage	2004
Bill Walsh	The Elephants of Style	2004
Pam Peters	The Cambridge Guide to English Usage	2004
June Casagrande	Grammar Snobs Are Great Big Meanies: A Guide ot Language for Fun and Spite	2006
	Grammar and Style	2007

Appendix B

Test Items

Item	Category	Handbook Repeats	form/ function	Rule Summary
aggravate	Lexical	14	form	*aggravate* means make worse, not irritate, exasperate, annoy
comprise	Lexical	17	form	whole *comprises* the parts
disinterested	Lexical	16	form	*disinterested* means impartial
infer	Lexical	17	form	A writer *implies* something, a reader *infers* it
nauseous	Lexical	16	form	*Nauseous* means causing nausea
adverse	Lexical	12	form	*adverse* means harmful or unfavorable; *averse* means having a feeling of opposition
between	Lexical	17	form	used with two and when the relationship is between individual items.
among	Lexical	17	form	used with more than two; a group treated as a collective unit
further	Lexical	18	form	abstract distance
farther	Lexical	18	form	literal distance
less	Lexical	18	form	mass noun
literally	Lexical	15	form	don't use in a figurative sense
different than	Lexical	18	function	*different* doesn't compare; use *from*
try and/ try to	Lexical	17	function	use *try to*
very unique	Lexical	16	function	*unique* is an absolute
brang	morphology	1	function	*brought* as past of *bring*
dived/dove	morphology	11	function	*Dived* is the past of *dive*
drunk	morphology	10	form	past tense: *drank*; past participle: *drunk*
hung/ hanged	morphology	13	function	criminals are *hanged*; posters are *hung*
these kind of	morphology	5	function	*kind* should be plural to agree with *these*

bad as intensifier	morphology	11	form	Avoid adjectival "bad" in an intensive sense; OK as an adverb
first/firstly	morphology	9	function	Don't use *firstly*; *first* is an adverb
real/really	morphology	12	function	*really* is an adjective
slow	morphology	10	form	adjective
myself	morphology	12	form	Use *myself* only as an intensive or reflexive
criteria	morphology	9	form	plural
data	morphology	16	form	plural
media	morphology	10	form	plural
has went	morphology	3	function	*has gone*
we was	morphology	14	function	*we were*
alot	Spelling	10	function	misspelling
all together	Spelling	12	Form	gathered in one place or all acting together
all ready	Spelling	12	form	completely prepared
already	Spelling	10	form	prior
alright	Spelling	18	function	misspelling
awhile	Spelling	13	form	an adverb meaning "for a short time"; cannot be the object of a preposition
a while	Spelling	13	form	A noun referring to a length of time
except	Spelling	11	form	Exclude
effect	Spelling	17	form	noun
affect	Spelling	17	form	verb
allude	Spelling	9	form	make an indirect reference (vs. elude)
discreet	Spelling	11	form	prudent, judicious
discrete	Spelling	11	form	separate, distinct
emigrate	Spelling	11	form	*emigrate* from (*immigrate* to)
principle	Spelling	14	form	a general truth or rule of conduct (vs. *principal*)
Between you and I (X and I as object)	Syntax	10	function	use object form
gerund with possessive	Syntax	13	function	use possessive before gerund, not accusative

Him and X as subject	Syntax	8	function	use subject form
that/which	Syntax	20	function	use *that* for restrictive clauses
who/whom	Syntax	20	function	use *whom* for objective
compound subject	Syntax	14	function	use plural verb
either is	Syntax	12	function	use singular verb
Feel bad/badly	Syntax	15	function	feel *bad* and not *badly*
multiple negation	Syntax	12	function	avoid
singular "they"	Syntax	4	function	use plural verb
only	Syntax	13	form	use immediately before the words it modifies
reason is because	Syntax	14	function	the reason takes a noun clause
split infinitive	Syntax	16	function	avoid
stranded preposition	Syntax	6	function	avoid
who/that	Syntax	20	function	*who* refers to animate antecedents

Appendix C

Excerpt from a Sample Survey Instrument

Note: There were thirty versions of this survey, each with twelve different sentences. The following sample survey contains only four sentences, all of which were taken from the actual study.

Demographic Questions:
Are you a native speaker of English?
- Yes
- No

How old are you?
- under 20
- 20 to 30
- 31 to 40
- 41 to 50
- 51 to 60
- 61 to 70
- 70+

What is the highest level of education you have completed?
- Some high school
- High school
- 1 year of college
- 2 years of college
- 3 years of college
- 4+ years of college
- Graduate degree

What is your gender?
- Female
- Male

Instructions:
In the following questions, you will read a sentence that has been published in a magazine or newspaper. Some sentences may contain an error that breaks a rule of written American English. Some sentences may contain no error. For each sentence, you will be asked to:

1. Identify whether or not you see an error.
2. If you see an error, identify how much that error bothers you on a scale from 1 to 6. A score of 1 means that you see the error but it doesn't bother you at all. A score of 6 means that you are extremely bothered by the error.
3. If you see an error, type which word(s) in which the error appears. DO NOT COPY AND PASTE ENTIRE SENTENCES.

1. Sentence: There's a difference between not talking and being utterly, truly silent.

- I see an error
- I see no error

If *I see an error* is selected:

If you see an error in the sentence above, how much does it bother you? (1 means not at all, 6 means very much)

- 1
- 2
- 3
- 4
- 5
- 6

Please type the word(s) that contain the error.

2. Sentence: We have voted into office people who we trust to do what we want for our community.

- I see an error
- I see no error

If *I see an error* is selected:

If you see an error in the sentence above, how much does it bother you? (1 means not at all, 6 means very much)

- 1
- 2
- 3
- 4
- 5
- 6

Please type the word(s) that contain the error.

3. Sentence: Here are some tactics that will help you catch them.

- I see an error
- I see no error

If *I see an error* is selected:

If you see an error in the sentence above, how much does it bother you? (1 means not at all, 6 means very much)

- 1
- 2
- 3
- 4
- 5
- 6

Please type the word(s) that contain the error.

4. Sentence: As marriage and parenthood have receded farther into the future, the first years after college have become, arguably, more carefree.

- I see an error
- I see no error

If I see an error is selected:

If you see an error in the sentence above, how much does it bother you? (1 means not at all, 6 means very much)

- 1
- 2
- 3
- 4
- 5
- 6

Please type the word(s) that contain the error.