Spring 2013

Plurality Cues and Non-Agreement in English Existentials

Robin Melnick
San Jose State University

Follow this and additional works at: https://scholarworks.sjsu.edu/etd_theses

Recommended Citation

Melnick, Robin, "Plurality Cues and Non-Agreement in English Existentials" (2013). *Master's Theses*. 4292.
DOI: <https://doi.org/10.31979/etd.fftf-c5k5>
https://scholarworks.sjsu.edu/etd_theses/4292

This Thesis is brought to you for free and open access by the Master's Theses and Graduate Research at SJSU ScholarWorks. It has been accepted for inclusion in Master's Theses by an authorized administrator of SJSU ScholarWorks. For more information, please contact scholarworks@sjsu.edu.

PLURALITY CUES AND NON-AGREEMENT IN ENGLISH EXISTENTIALS

A Thesis

Presented to

The Faculty of the Department of Linguistics and Language Development

San José State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

by

Robin Melnick

May 2013

© 2013

Robin Melnick

ALL RIGHTS RESERVED

The Designated Thesis Committee Approves the Thesis Titled
PLURALITY CUES AND NON-AGREEMENT IN ENGLISH EXISTENTIALS

by

Robin Melnick

APPROVED FOR THE DEPARTMENT OF
LINGUISTICS AND LANGUAGE DEVELOPMENT

SAN JOSÉ STATE UNIVERSITY

May 2013

Dr. Soteria Svorou	Department of Linguistics and Language Development
Dr. Stefan Frazier	Department of Linguistics and Language Development
Dr. Kevin Moore	Department of Linguistics and Language Development

ABSTRACT

PLURALITY CUES AND NON-AGREEMENT IN ENGLISH EXISTENTIALS

by Robin Melnick

This paper furthers the discussion of variable agreement in English existential constructions. Previous studies across dialects have shown that *there+be* with a plural notional post-copular subject is frequently realized with contracted singular agreement, for example, “There’s many articles on this topic.” Prior work in building probabilistic models for predicting the presence of agreement or non-agreement in any given such *there+be* sentential context has investigated a variety of factors with potential influence on this variation, but the present study provides evidence for the inclusion of two novel and significantly predictive elements: a plurality “cue distance” and a new taxonomy for determiner type. The latter references each form’s strength in terms of number semantics, rather than along the lines of definiteness employed in traditional determiner classifications. These new factors are, in turn, motivated by a general formulation, the *Weak Number Hypothesis*, which offers further insight into factor significances found by prior works. Multiple corpus studies and logistic regression model analysis provide empirical support for the central hypothesis and its attendant predictions.

ACKNOWLEDGMENTS

I am deeply grateful to the faculty of the Department of Linguistics and Language Development at San José State University for nurturing a newfound academic passion in someone coming from so many years away from the academy. In particular, I would like to thank the members of my thesis committee, Professor Stefan Frazier, Dr. Kevin Moore, and thesis advisor, Professor Soteria Svorou, for their support of the present work.

Thanks to Dr. John Fry for his feedback on the project in its early stages. Special and affectionate gratitude goes to both Professor Svorou and to Professor Manjari Ohala, the mentors who first and most enthusiastically encouraged me to come to SJSU. I would not be where I am today without them.

I would like to further acknowledge the support and encouragement of Professor Tom Wasow and Professor Emeritus Joan Bresnan of Stanford University. Their work in probabilistic syntactic variation provided the impetus to take my research in this direction.

Finally, I am, as always, most deeply grateful to my family for their unswerving love and support.

TABLE OF CONTENTS

List of Figures	viii
List of Tables	ix
Chapter 1 – Introduction	1
Chapter 2 – Background	9
Agreement	9
English Existential <i>There+Be</i>	12
Probabilistic Syntactic Variation	16
Statistical Procedures	18
Chi-square test	18
Regression analysis	20
Binary logistic multiple regression	21
Mixed-effects modeling	22
Chapter 3 – Study 1: Distributional Analysis	24
Data	25
Determiner Classification	26
Cue Distance	29
Other Measures	30
Hypotheses	32
Results	32
Overall agreement	32
Noun distance	33
Cue distance	34
Determiner type	34
Determiner type and cue distance	35
Chapter 4 – Study 2: Multi-Corpus Investigation	38
Switchboard	38
Overall agreement	38
Noun and cue distances	38
Determiner type	39
Determiner type and cue distance	40
Charlotte and Callhome Corpora	41
Overall agreement	42
Determiner type and cue distance	42
Discussion	43

Chapter 5 – Study 3: Modeling the Effects on Discord for New Factors	46
Model and Coding	46
Modeling Cue Distance	50
Modeling Determiner Type	51
Chapter 6 – Discussion	52
Chapter 7 – Conclusion	56
References	58
Appendix A – Corpus Sample	62

LIST OF FIGURES

Figure 1.	Rates of discord for Cue Distance and Noun Distance against sample average.	34
Figure 2.	Percentage of non-agreement by type of determiner.	35
Figure 3.	As predicted by the WNH, determiner-type relative discord percentages align closely with associated Cue Deltas.	36
Figure 4.	Rates of discord for Cue Distance and Noun Distance in the Switchboard sample.	39
Figure 5.	Non-agreement by type of determiner, Switchboard sample.	40
Figure 6.	Cue Delta alignment with determiner-type discord percentages for Switchboard sample.	41
Figure 7.	Comparative discord rates and corresponding average Cue Delta values for the four corpora.	43

LIST OF TABLES

Table 1.	Factors in Riordan (2007).	3
Table 2.	Milsark (1977) determiner classification.	7
Table 3.	Frequencies of existential types by NP plurality (Switchboard sample). ...	25
Table 4.	Huddleston and Pullum (2002) non-count-quantificational nouns.	27
Table 5.	Determiner-type factor coding.	28
Table 6.	Factors coded for analysis of determiner-type and “cue distance” effects.	31
Table 7.	Rates of discord by existential form (with plural NPs).	33
Table 8.	Reproduction of Riordan (2007) mixed-effects model fitted to Study 1 sample (N = 500).	48
Table 9.	Classification accuracy on Study 1 sample (N = 500).	49
Table 10.	Mixed-effects model updated to include Cue Distance factor.	50
Table 11.	Classification accuracy with Cue Distance added to regression.	51

Chapter One – Introduction

Queen: There's two of you; the devil make a third...

Shakespeare, *Henry VI, Part 2* (III.ii)

Trinculo: They say there's but five upon this isle: we are three of them; if th' other two be brain'd like us, the state totters.

Shakespeare, *The Tempest* (III.ii)

Shakespeare famously scripted his works to appeal to audiences both high and low, in part by reflecting their respective language traits in the speech of representative characters. That haughty Queen Margaret and the fool Trinculo each employs an existential construction exhibiting a lack of notional subject-verb agreement offers a first hint then that non-agreeing *there+be* is neither a new phenomenon nor one limited to non-prestige dialects. On the latter point, more robust evidence perhaps comes from corpus studies by Crawford (2005) that dispute the notion that non-agreement with *there+be* is limited to casual discourse or reflective of a lack of education. On the historical element, Meechan and Foley (1994, citing Quirk and Wrenn's 1957 study of Old English) note that existential verbs were realized with variable agreement as far back as 1000 A.D., for example:

- (1) δār sceal beon gedrync and plega
 there must.1/3SG be drinking and merrymaking

(Quirk & Wrenn, 1957:76)

The phenomenon has been variously termed “singular agreement” (Meechan & Foley, 1994), “non-agreement” (Schütze, 1999), “non-concord” (Martinez Insua & Palacios Martinez, 2003), “disagreement,” “discord,” and indeed “non-prestige” (Riordan, 2007). Likewise, researchers have approached the discussion from a variety of linguistic perspectives: syntax theoretic (Chomsky, 1995; Meechan & Foley, 1994; Milsark, 1977; Pietsch, 2005; Rupp 2005; Sauerland & Elbourne, 2002; Schütze, 1999; *inter alia*); historical (Breivik, 1990; Breivik & Martinez Insua, 2008; Hay & Schreier, 2004); sociolinguistic (Britain, 2002; Britain & Sudbury, 2002; Crawford, 2005; Schilling-Estes & Wolfram, 1994; Tagliamonte, 1998); and most recently, probabilistic variation based on multivariate logistic regression analysis (Riordan, 2007).

Collectively, these studies have identified a wide range of factors associated to one degree or another with numerical discord in English existential constructions. In the most statistically thorough investigation, Riordan (2007) lists no fewer than 27 factors used across prior corpus work with *there+be* before proceeding to annotate his own data for ten of these, including linguistic, processing, social, and discourse elements (see Table 1).

Table 1. Factors in Riordan (2007).

<u>Linguistic</u>	
	Type of determiner
	Sentential polarity
	Plural head noun ends in –s
<u>Processing</u>	
	Distance between copula and head noun of post-copular NP
	Disfluency between copula and head noun
	Weight of post-NP sequence
<u>Discourse</u>	
	Primary discourse mode
<u>Social</u>	
	Speaker age
	Academic role ¹
	Gender

Crucially for the present work, several of the prior quantitative multivariate corpus studies have explored the connection between discord and the distance from the copula to the head of the post-copular NP. The results, however, have been mixed: While Tagliamonte (1998) and Britain and Sudbury (2002) found the effect to be statistically significant, Meechan and Foley (1994) and Hay and Schreier (2004) did not. Riordan (2007) found this distance to be significant only in combination with determiner type. The latter element provides a second crucial component for the present work. Each of the aforementioned studies identifies determiner type as being at least somewhat correlated with *there+be* agreement variation, but while the association itself is well-documented,

¹ Riordan worked with the Michigan Corpus of Academic Spoken English (Simpson et al., 2002), in which all speakers have one role or another within a university setting.

explanations for why this should be the case have been limited and largely speculative.

Like Riordan (2007), the current investigation works with measures of distance and determiner type in combination, but the central, novel insight of the present study arises in taking a different perspective on what is most salient about these elements. Here, a new intuition about *there+be* discord is that processing plays an essential role, agreeing with Cheshire (1999) that having a limited capacity for “look ahead,” speakers are frequently led by discourse pressures to select the singular, contracted *there’s* as what Breivik and Martinez Insua (2008) call a “presentative formula.” The new insight, though, is that while prescriptive agreement is, of course, between copula and post-copular notional subject, it is not all the way to the head noun itself that speakers must necessarily look in their planning; rather, it is to whatever post-copular element it is that provides the first unambiguous signal as to the grammatical number of the NP. Crucially, the semantics of an intervening determiner may provide such an advance signal as to the ultimate plural or singular grammatical number of the NP, regardless of further surface distance to the head noun.

While this seems intuitive, prior studies have focused on head-noun distance, not “cue distance,” and considered and classified determiners largely based on definiteness, not number semantics. This leads to the present work’s central proposal:

(2) **Weak Number Hypothesis**

(with regard to post-copular agreement contexts)

When present between copula and post-copular head noun, additional elements that are semantically weak in terms of number signaling contribute to discord, *ceteris paribus*. Intervening elements that provide strong plurality cues tend to promote concord.

The *Weak Number Hypothesis* (henceforth WNH) has the following empirically testable implications:

- (3) i. Let us define a novel metric, *Cue Distance*, as the number of words from copula to the first temporal element that provides an unambiguous signal with regard to the plurality of the associated post-copular NP. The WNH predicts that such a Cue Distance will correlate with discord in post-copular agreement contexts, and specifically, it will have a more significant effect than the overall distance to the head noun.
- ii. The rate of discord associated with a particular type of determiner should correlate with how much further beyond the determiner a speaker must plan in order to be certain of NP number. Determiners that are strongest in terms of their number semantics—those that are unambiguously compatible with either only plural or only singular head nouns—should correlate with lower rates of discord. Number-ambiguous determiners should correlate with higher rates of discord.

The prediction of (3.i) is that with distance to head noun held constant, an intervening plurality signal—meaning a shorter “cue distance” (CD)—should promote agreement. For example, the WNH predicts that agreement is more likely in (4.e) than in (4.a). Though the head nouns are identical (*books*) and at the same total distance (five words), Cue Distance in (4.e) is substantially less, given the presence of the strongly plural quantifier *several*. (Underlining indicates the first unambiguous number cue.)

- (4) a. *There+be* **excellent**, interesting and insightful **books** on the subject. (CD=5)
- b. *There+be* **excellent**, interesting and insightful **info** on the subject. (CD=5)
- c. *There+be* **some** interesting and insightful **books** on the subject. (CD=5)
- d. *There+be* **some** interesting and insightful **info** on the subject. (CD=5)
- e. *There+be* **several** interesting and insightful **books** on the subject. (CD=1)

In contrast to (4.e)—for which intervening plural-only determiner *several* yields a Cue Distance of one word, just the determiner itself—in (4.a) and (4.b) the speaker must look ahead all the way to *books* and *info*, respectively—a total Cue Distance of five words in each case—to know whether singular or plural agreement is required.

The second prediction of the WNH, (3.ii), is that agreement is also more likely in (4.e) than in (4.c). Here not only are the head nouns the same and at the same distance, but each of these examples also has an intervening determiner immediately following the copula. Moreover, the determiner taxonomies employed in prior studies would identify no distinction here; the determiners *some* and *several* would each be classified as a similarly indefinite quantifier. However, while *several* in (4.e) is strongly plural—

compatible only with a plural head noun—*some* in (4.d) is semantically weak in terms of number signaling, offering no clue as to the grammatical number of its associate. This is seen in the contrast between (4.c) and (4.d), both of which have the same determiner, *some*, but with head nouns *books* (plural) and *info* (grammatically singular), respectively.

This brings us back to the second novel insight of the present work: most discussions of determiner “strength”—for example, the taxonomy articulated by Milsark (1977), variations of which are employed by several of the aforementioned studies—are concerned primarily with definiteness (see Table 2).

Table 2. Milsark (1977) determiner classification.

<u>Weak Determiners</u>	<u>Strong Determiners</u>
<i>a</i>	definites: <i>the</i> , demonstratives, pronouns, possessives
number determiners	universals: <i>all</i> , <i>every</i> , <i>each</i> , <i>any</i> (when not a polarity item of <i>some</i>)
<i>some</i> , <i>many</i> , <i>few</i> , <i>lots of</i> , <i>plenty of</i> , <i>several</i>	<i>some (of the)</i> , <i>many (of the)</i> , <i>most</i>
Ø plural and mass determiners in their non-universal reading	Ø determiner in universal reading
<i>no</i>	

In Milsark’s taxonomy, *some* and *several* are both “weak” determiners because they are equally indefinite. Conversely, *the* and *these* are both “strong” because they are equally definite. For the purpose of predicting agreement in post-copular “look-ahead” contexts, however, the hypothesis of the present work is that relative strength of

definiteness is irrelevant; it is the relative semantic strength of number signaling that is expected to influence agreement. (An alternative taxonomy along these lines is detailed in Study 1, Chapter 3.) Unambiguously plural determiners such as *several* (Milsark: “weak”) and *these / those* (Milsark: “strong”) as well as unambiguously singular forms such as *much* and *this / that* are all expected to promote number concord, while grammatical-number-ambiguous forms such as *some* and *the* are expected to equally promote discord, even though each of these pairings contrasts in terms of definiteness.

The primary aim of the present work then is to explore empirical support for the *Weak Number Hypothesis* via a series of corpus-based studies. We begin with a bit of background (Chapter 2), including brief sketches of previous work in non-agreement, English existential constructions, and probabilistic syntactic variation, as well as a quick introduction to the statistical procedures used in later sections. Next, Study 1 (Chapter 3) documents an initial distributional analysis using the Michigan Corpus of Academic Spoken English (Simpson et al., 2002). Study 2 (Chapter 4) investigates the robustness of the observed effects by seeking corroboration in three other spoken corpora: Switchboard (Godfrey et al., 1992), Callhome (Kingsbury et al., 1997), and the Charlotte Narrative and Conversation Collection (Charlotte, via the American National Corpus, Reppen et al., 2005). Study 3 (Chapter 5) further tests the agreement-predicting significance of Cue Distance and determiner number semantics by adding these new factors to an existing multivariate predictive model for this phenomenon (Riordan, 2007). The concluding discussion further sets the paper’s central hypothesis in its theoretical context and suggests several possibilities for follow-on work.

Chapter Two – Background

The present study situates itself at the intersection of prior linguistic work on: subject-verb agreement, in general; non-agreement in English existential constructions, in particular; and syntactic alternation—as in, for example, the English dative construction—studied via multivariate probabilistic models of speaker choice.

Agreement

Several factors not specific to English *there+be* constructions may affect the production of subject-verb agreement. First, structural priming—also known as syntactic persistence—refers to the tendency for either the production or comprehension of a particular form to increase the probability of this form’s later production. On this account, a speaker having just uttered (5.a) is incrementally more likely to produce the discordant continuation in (5.b).

- (5) a. There’s one train this morning, but ...
 b. ... there’s two more this afternoon.

Bock (1986) attributes this tendency to an increase in the level of mental activation of a particular form via its first use, whether as speaker or comprehender, making a form “primed” and therefore more easily accessed for later production. In Bock’s experiments, participants were first given an initial priming sentence to read aloud then were asked to describe an unrelated picture. Evidence from her studies suggests that the effects of priming are independent of sentence content. In following work, Bock and

Loebell (1990) found that priming effects can be isolated more specifically within syntax. The authors employed materials that removed any overlap in meaning or sharing of lexical items across the initial priming utterance and later related production. Their conclusion is that it is prior activation of a like hierarchical constituent structure that is responsible for Bock's previously observed priming effect, not simply a repetition of a similar semantic proposition.

While the priming work was taken as applying to all syntactic production, non-agreement phenomena among others, Bock and Miller (1991) subsequently delved further into non-agreement specifically, looking at locality effects, wherein a differently numbered NP interceding between head noun and verb tends to “attract” agreement, for example in (6).

(6) The cost_(SG) of the improvements_(PL) have_(PL) not yet been estimated.

In elicited sentence-completion experiments, the researchers found that varying the plurality of the immediately pre-verbal (local) noun had a large and reliable effect, while varying subject-like semantic features, for example, animacy, had little effect. The authors take these findings as supporting explanations of language production as segregated into relatively autonomous components.

Attempting to tease out which linguistic processes exert most control over agreement, Bock, Nicol, and Cutting (1999) worked with collective nouns—for example, *cast*, *team*, *jury*—which exhibit a contrast of notional plurality with singular grammatical number. In another elicited-continuation task, participants completed sentences where

some items tested verb agreement, as in (7.a), while others checked anaphoric pronoun agreement, as in (7.b).

- (7) a. The cast in the soap operas was / were popular.
 b. The cast in the soap operas watched itself / themselves. (Bock et al., 1999:334)

The researchers found that verb agreement in these conditions showed greater influence of grammatical number, while pronominal agreement exhibited greater influence of notional number, concluding that this argues against efforts to anchor linguistic number and number agreement in a single semantic or linguistic source (Bock et al., 1999:343).

Looking at a theoretical approach to agreement—within the Minimalist Program (MP), specifically—den Dikken (2001) does not actually consider non-agreement to be a matter of variation, but rather treats *jury / team*-type collective nouns—which he terms “pluringulars”—as uniformly calling for singular agreement in certain contexts, plural agreement in others. In particular, the author suggests that collectives always take singular agreement in existential *there+be* constructions, as in (8.b).

- (8) a. The committee (is | are) undecided.
 b. (There’s | *There are) the committee now.

In yet another sentence-completion task with collectives, Humphries and Bock (2005) used minimally varying prompts to bias nominals for either singular (collected) or plural (distributed) construal. For example, considering collective noun *gang*, the single-

word difference in (9) between prepositions *on* vs. *near* was found to bias for a distributed (plural) vs. collected (singular) reading, respectively.

- (9) a. The gang on the motorcycles (BIAS FOR PLURAL)
 b. The gang near the motorcycles (BIAS FOR SINGULAR)

Finally, Bock, Eberhard, Cutting, Meyer, and Schriefers (2001) returned to locality effects (i.e., agreement “attraction”) with experiments using words that provide a range of permutations of grammatical and semantic number. As the authors note, words such as *suds* are plural both grammatically and notionally; a collective such as *army* can be construed as semantically plural despite grammatical singularity; and finally, pluralia tantum—words such as *scissors* which appear only in the grammatical plural—are nevertheless (arguably) semantically singular. Working in both English and Dutch, the researchers used these word types as preverbal distractors in a judgment task, asking participants to consider examples such as (10).

- (10) If you were thinking about the advertisement for the scissors, would you be thinking about one thing or more than one thing?

(Bock et al., 1999:96)

The results suggest that only grammatical number increased attraction, not the distractors’ number meanings.

English Existential *There+Be*

Turning more specifically to the present study’s target construction, linguistic researchers have, as previously noted, approached non-agreement with English

existentials from a variety of angles: from syntax theoretic modeling to the tracing of *there+be*'s historical development. Most salient for the present work are the prior studies from a variationist perspective, looking at non-agreement not simply as performance error, in putative violation of a supposedly fixed grammar, but rather as one of a set of alternative productions and, in fact, potentially predictable based on a range of factors.

The majority of these variationist treatments have explored the matter as a sociolinguistic phenomenon, focused on differences among groups, whether drawn along the lines of geography/dialect, age, gender, or education. Meechan and Foley (1994), for example, initially note that earlier work had found determiner type to be a potential influence on *there+be* non-agreement—albeit in terms of a strong/weak, definiteness-based determiner hierarchy, not the number-signaling, semantics-based determiner typology proposed by the present work—but their own subsequent corpus analysis finds no significant effect of determiner type. What they do find important instead is level of education, that is, a social—as opposed to linguistic—factor. Crawford (2005), working along another dimension, employed a large multi-corpus study to demonstrate that non-agreement in *there+be* is not simply, as others had suggested, an artifact of informal conversation, but rather that it is well-evinced in a variety of registers, including both written and spoken monologue. Other studies are even more purely socio-focused, for example Schilling-Estes and Wolfram (1994), wherein the authors identify existentials as one of a handful of constructions that demonstrate the *was* / *were*-leveling characteristic of Ocracoke (North Carolina) Vernacular. The authors suggest that in positive-polarity past tense, Ocracoke residents near-uniformly employ *there was*, regardless of NP

plurality, while in past tense with negative polarity, these speakers near-uniformly produce *there weren't*. Interestingly, Britain (2002) finds essentially the same polarity-driven *was / were*-leveling among the Fenland dialects of England's East Midlands region.

Another group of studies similarly finds its primary motivation in explorations of between-groups sociolinguistic variation, but this latter body of work goes beyond socio factors—dialect, age, gender, education, and so on—in recognizing that *there+be* agreement variation also appears to be subject to within-speaker variation based on a range of more purely linguistic factors. While some of these studies seem to merely control for within-speaker effects so as to better focus on between-groups variation, others more fully embrace the simultaneous study of variation along multiple dimensions. In the latter camp, we find Tagliamonte (1998). Her diachronic study considers the increasing use of past-tense non-standard *was* in York (northeast England), with existential *there+be* just one of the environments in which she observes this trend. While her primary focus is the social pattern of change, identifying this as a case of female-led innovation, she also explores such linguistic factors as polarity, grammatical person, distance to post-copular subject NP, and type of NP, as well as lexical influences. Like the earlier Meechan and Foley (1994) study, Tagliamonte also includes determiner type, but once again based on a definiteness-oriented—rather than number-signaling—hierarchy. Here, she finds as others had that determiner type, at least as thus encoded, is not a significant factor in the variation. Britain and Sudbury (2002), in turn, model a similar set of factors in their cross-dialectal study of existential *there+be* variation. The

authors find parallel patterning of factors across a pair of comparatively isolated English dialects, from New Zealand and the Falkland Islands, respectively.

Though of slightly less interest to the present work, it is worth noting that the issue of *there+be* non-agreement has also been taken up by historical linguists and theorists, as well. Among the historically minded, Breivik and Martinez-Insua (2008) explore the phenomenon as an example of grammaticalization, while Hay and Schreier (2004) trace the rise, decline, and subsequent rise again of *there+be* non-agreement in New Zealand. Others, such as Breivik (1990) and Pietsch (2005), trace the construction's history largely as support for their proposals of theoretical treatments for uniform singular agreement.

Lastly, there are a smaller number of prior works on English existential *there+be* that, like the present study, concern themselves exclusively with the exploration of factors underlying within-speaker—rather than between-groups—variation. Martinez-Insua and Palacios-Martinez (2003), working with a corpus composed solely of modern standard British English, explore a wide range of linguistic factors suggested by earlier studies as influencing *there+be* agreement variation, coming away with the further observation that coordinated post-copular NPs—as in (11), for example—are particularly subject to non-agreement.

(11) There's a boy and a girl in the yard.

The most direct corollary to the present work, finally, is Riordan (2007), wherein the author annotated a large, modern corpus of spoken standard American English for a

total of ten linguistic, social, and discourse elements, concluding, much like Crawford (2005), that register plays a key role in *there+be* variation. Further aspects of this work are detailed in the empirical studies below, particularly highlighting those areas where the present work's approach and results each contrast with the earlier study.

Probabilistic Syntactic Variation

The tradition of looking at linguistic variation as something beyond the realm of pure dialectology dates at least to Labov (1964), if not beyond, and as seen above, succeeding generations of sociolinguists have not only expanded their sets of sociological variables, but they have sought to account for a number of possible sources of within-speaker variation, as well. More recently, in fact, another body of work has arisen to focus squarely on within-speaker variation, especially in syntax, from a purely linguistic perspective. Whereas traditional mainstream syntactic theorizing has held that variation is explicitly an extra-linguistic matter, this newer line of research concerns itself exactly with investigation of how a given speaker—meaning that sociological variables are held constant—might come to choose between multiple, equally grammatical syntactic forms that express the same semantic content. For example, while theoreticians may debate the most appropriate grammar for separately licensing the pair of competing constructions that form the English dative alternation—as in the semantically equivalent sentences in (12)—they are generally not concerned with exploring the conditions that might influence a speaker's choice between such alternatives.

- (12) a. Pat gave [a book]_{NP-THEME} [to a student]_{PP-RECIPIENT}.
 b. Pat gave [a student]_{NP-RECIPIENT} [a book]_{NP-THEME}.

The approach generally taken by syntactic variationists is to treat such alternations as probabilistic, which is to say that such a choice may be predictable (if certainly imperfectly so) from the set of relative likelihoods contributed by multiple contextual factors. For example, Bresnan, Cueni, Nikitina, and Baayen (2007), working with the dative alternation, find that as the length of RECIPIENT increases relative to the length of THEME, it tends to promote the prepositional-RECIPIENT option (a); a pronominal RECIPIENT tends to promote the NP-RECIPIENT option (b); relative definiteness of the THEME tends to promote the prepositional-RECIPIENT option (a); and so on for a set of more than a dozen such factors. Further, each such tendency is assigned a weighted probability, a measure of the strength of each factor's influence relative to the others. By contrast, traditional grammars make only categorical statements in this regard, if any statement at all. For example, it is usually asserted that a pronominal RECIPIENT will always be realized as an NP indirect object, as in (12.b), never as an oblique argument (the object of a preposition) as in (12.a). In a multivariate probabilistic treatment, on the other hand, pronominality of RECIPIENT remains strongly weighted relative to other factors, but under the right conditions the model can still predict the prepositional-RECIPIENT with some relative likelihood. This probabilistic modeling approach has been applied to a wide range of syntactic alternations, including just for example: English verb-particle placement (Gries, 2003); English complementizer ("that") optionality (Jaeger, 2006); and optional "that" with English relative clauses (Wasow et al.,

2011). See (13-15), respectively.

(13) a. Dale picked up the clothes off the floor.

b. Dale picked the clothes up off the floor.

PARTICLE SHIFT

(14) a. Francis knows that I'll be there.

b. Francis knows I'll be there.

OPTIONAL COMPLEMENTIZER

(15) a. ... the book that you're reading.

b. ... the book you're reading.

OPTIONAL RELATIVIZER

In the present work, Study 3 involves developing such a probabilistic model for agreement variation with English existential *there+be*. The statistical method used to build these predictive models is binary logistic multiple regression, a technique which will be outlined briefly in the following section.

Statistical Procedures

The studies included in the present work employ two principal statistical measures. A brief primer on these metrics is presented.

Chi-square test. Shorthand for “Pearson’s chi-square test” and usually abbreviated as χ^2 , chi-square measures the extent to which an observed set of values matches expected results for a given hypothesis. This is often referred to as measuring the “goodness of fit” of observed to expected, and is only appropriate for categorical data, meaning things that can be counted. A statement of results typically presents the value for the chi-square test itself (χ^2), the “degrees of freedom” (*df*), and a corresponding “*p*-value” as an indicator of the significance of the result. A higher χ^2 indicates less

goodness of fit, so if the results are described as significant, this is to say that observations significantly differ from expectations. *Degrees of freedom* is a measure of the complexity of the comparison, equal to one less than the number of categories being compared. Intuitively, the greater the *df*, the greater our tolerance for a difference between observed and expected while still considering it to be a good match. Finally, *p* will be a value between 0.00 and 1.00, where the lower the *p*-value, the greater the significance of the difference between observations and expectations. By convention, a *p*-value of less than or equal to 0.05 is considered to be significant.

What does a *p*-value of 0.05 actually signify? We understand that there is always a certain amount of randomness in the sampling of data so we want to consider the possibility that a difference between two sets of data might not reflect a “true” difference. Suppose all we were doing was drawing a random sample from one larger body of data. Here we generally expect the full data and a sample drawn from it to have similar distribution characteristics, but for a given random sample there is some chance of it exhibiting a somewhat larger difference, again due just to the nature of random sampling. A *p*-value of 0.05 would indicate that only once in every 20 random samplings would we expect to find the observed level of difference due to sampling error alone. This would therefore be considered likely to be evidence of a “real” difference between the groups of data, not one due to the randomness of sampling alone. With a lower *p*, generated by a larger difference yielding a greater χ^2 value, it would be considered even less likely that the observed difference is due simply to sampling. For example, $p = 0.01$ would suggest that only once in 100 random samplings from a larger group of data would we be likely

to draw a sample exhibiting this much difference. In the other direction, as p increases, it becomes more and more likely that we could encounter such a (smaller) difference (lower χ^2) just in a random sample from a larger body of data, leaving us less and less confident that we are observing a “true” difference and not just sampling error.

Regression analysis. Consider this by way of analogy. Suppose that every human baby weighs exactly five pounds at birth and that we subsequently gain exactly five pounds for each year of life.² If so, we could calculate any person’s exact weight by the simple linear equation in (16).

$$(16) \text{ weight} = 5 + (5 * \text{age})$$

Of course people gain and lose weight in a highly individual manner throughout life. Still, suppose we had a set of observed actual weights and ages for a large sample of people and from these we wanted to come up with a simple linear equation, of the same general form as (16), that would give us as good an *estimate* of an individual’s weight as possible given the limiting simplicity of the formula. This equation might take the form of (17).

$$(17) \ y = \beta_0 + \beta_1 * x$$

Here, β_0 is our estimate of average birth weight, β_1 is our estimate of average weight gain for each year of life, x is age, and y is our resulting estimated weight.

Statistically, regression analysis is the procedure, offered as a built-in function within a number of statistics software packages, that we use to generate our values (called

² This is certainly not a terribly realistic supposition, but it will make for a simpler illustration.

Estimated Coefficients, or Estimates) for β_0 and β_1 . The goal of regression analysis is to yield estimated coefficients that produce a set of y values as closely matched to the actual weights in our population sample as possible, given the limiting simplicity of our “model,” that is, our linear expression.

Now suppose we want to add variables to our linear formula in order to yield better estimates, perhaps including some measure of how much one exercises plus whether or not one smokes. Our formula is now akin to (18).

$$(18) \ y = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \beta_3 * x_3$$

Here, x_1 is age; x_2 is exercise level; and x_3 is whether or not one smokes (yes/no, probably actually implemented as 1/0). A *multivariate* (or *multiple*) regression analysis yields a set of estimated coefficients, $\beta_0 \dots \beta_3$, calculated to yield estimated weights as close as possible to the actual observations.

Binary logistic multiple regression. Finally, instead of predicting weight as the output of our model, suppose we want to yield a yes/no estimate of life expectancy, simply whether or not (yes/no, or 1/0) we expect one to be alive at age x_1 , given exercise level x_2 , and whether or not the person smokes (or smoked) x_3 . We now have a direct analog for the procedure employed in Study 3 of the present work. The (binary) predicted outcome in the present work is agreement vs. non-agreement; the x_i 's are the several linguistic and social factors present for any given sentence; and the β_i 's are the estimates yielded by regression analysis for the observed data culled from corpus analysis. For each sentence, we have the actual corpus observation for agreement/non-agreement,

and the model gives us a corresponding prediction for agreement. The quality, or “goodness of fit,” of our model will be a measure of how closely the predicted values match what is actually in the corpus. Most importantly, we can now use elements of this model to generalize across the language; the model’s β_i ’s are our estimates of the relative weighted influence of each linguistic factor in the within-speaker production decision involved in yielding *there+be* agreement or non-agreement.

Mixed-effects modeling. Finally, consider once more our weight-modeling analogy. Though real people gain and lose weight at their own individual rates, we have settled on a generalized estimate for a uniform rate of weight gain for all. This sounds worse than it is. In our actual model in Study 3, this corresponds to coming up with, for example, an estimate of the degree to which agreement is influenced by the distance between copula and post-copular agreement controller (NP). By generalizing across all speakers, we are able to make a broader statement about the nature of this effect in our common language (or dialect). In regression modeling, this is known as a “fixed effect.” Returning to our hypothetical weight model, we also assume that all babies have the same birth weight. Now, we might like to allow that each individual could have a distinct starting weight, while still leaving most factors as generalized, or “fixed.” In (18), this would correspond to estimating a differentiated β_0 for each individual in our data set. In regression analysis, this is known as a “random effect,” and employing both fixed and random effects is called “mixed-effects” regression. In Study 3 below, the agreement model is in fact incrementally improved by allowing that—aside from the influences of contextual factors—each individual speaker may have a distinct baseline rate for non-

agreement. Some speakers might uniformly employ singular agreement; some may produce 100% prescriptive agreement; and most likely fall somewhere along a gradient between these end points. Because our source corpus in this case came already tagged with a speaker ID for each utterance, we can thus model these varying baselines using mixed-effects regression with a random effect for speaker ID.

Chapter 3 – Study 1: Distributional Analysis

Empirical exploration of the *Weak Number Hypothesis* begins with a look at environments that favor non-agreement. To limit scope and better facilitate comparison with earlier studies, the current work focuses on agreement in *there+be* existentials in the present tense and with a plural post-copular associated NP.

While Tagliamonte (1998) worked exclusively with past-tense *was/were* variability in her York (U.K.) study, Riordan (2007) notes that most prior studies have found *there+be* agreement variability to be significantly more prominent in the present tense (cf. Crawford, 2005; Martinez Insua & Palacios Martinez, 2003; Meechan & Foley, 1994).

With regard to a focus on plural NPs, the present data concur with the consensus of prior studies that *there+be* discord is a near-categorically one-way phenomenon; plural NPs frequently present with the singular copula, but the reverse—singular NP with plural copula—is extremely rare. For example, in a 516-token sample of present-tense *there+be* constructions drawn at random from the three-million-word Switchboard corpus, 61.8% (128/207) of plural post-copular NPs presented with singular forms of the copula (*there+is/'s*), while just 0.3% (1/309) of singular post-copular NPs were associated with a plural form of the copula (*there+are/'re*). (See Table 3.)

Table 3. Frequencies of existential types by NP plurality (Switchboard sample).

<i>there+be</i> form					
	<u>Singular</u>		<u>Plural</u>		
	<i>there's</i>	<i>there is</i>	<i>there're</i>	<i>there are</i>	<u>total</u>
singular NP	267	41	0	1	309
plural NP	121	7	2	77	207
<u>total</u>	388	48	2	78	516
discord with singular NP:					0.3%
discord with plural NP:					61.8%

Data

We turn first to the Michigan Corpus of Academic Spoken English (MICASE). MICASE is a project of the University of Michigan, with a growing archive of transcripts totaling 1.85 million words at the time it was accessed for the present study. “Academic” speech in this context is not limited to “scholarly discussion,” but rather a wide variety of “speech which occurs in academic settings,” including “jokes, confessions, and personal anecdotes, as well as definitions, explanations, and intellectual justifications” (English Language Institute, 2002).

The MICASE web-based search interface was used to gather all sentences with present-tense forms of the English existential *there+be* construction: *there is*, *there's*, *there are*, and *there're*. From this database, a 500-token random sample was then extracted for annotation. (This sample is included in its entirety in Appendix A.) Negation was tokenized separately, meaning *there isn't* / *aren't* was transformed to *there is* / *are n't* before sampling. Annotation was performed by a combination of custom software and hand coding. Tokens from non-native speakers were excluded, as were uses

such as those in (19), including locative *there*—those only incidentally followed by a form of the copula—as well as past and future-tense forms such as *there's been* or *there are going to be*.

(19) a.	What they're gonna see <u>there is</u> an issue of adjustment ...
(ADV700JU047:S2) ³	LOCATIVE
b.	<u>There's been</u> no empirical research supporting it.
(LEL185SU066:S1)	PAST
c.	<u>There's going</u> to be all these rings ...
(LAB175SU033:SU-f)	FUTURE

Determiner Classification

As previously noted, the Milsark (1977) determiner taxonomy is inadequate for the present purpose because semantic plurality is orthogonal to Milsark's distinction along the lines of definiteness. A number of prior studies (Britain & Sudbury, 2002; Hay & Schreier, 2004; Meechan & Foley, 1994; Tagliamonte, 1998) do implement somewhat more granular distinctions than simple weak/strong, including separately calling out, for example, *no* or cardinal numbers. Tagliamonte, in particular, culls out partitives (e.g., *a lot of*, *a bunch of*). Interestingly, Riordan (2007) discusses the very fine-grained distinctions outlined in Huddleston and Pullum (2002), which in fact notes exactly the kind of semantic distinction proposed in the present study, that is, the extent to which certain forms are compatible with singular complements, plurals, or both (see Table 4). When Riordan proceeds to lay out the particulars of his own determiner encoding scheme,

³ The code to the left of the colon is the MICASE transcript ID. To the right of the colon is the individual speaker ID within the particular transcript.

however, he regroups Huddleston and Pullum’s three different types of “non-count quantificational nouns” (NCQNs) back into a single “NCQN” category and loses their number-semantics distinction.

Table 4. Huddleston and Pullum (2002) non-count-quantificational nouns.

Number-transparent quantificational nouns

*lot, plenty, lots, bags, heaps, loads, oodles, stacks,
remainder, rest, number, couple*

Non-count quantificational nouns selecting a singular oblique

*a great/good deal of, a smidgen/bit of, a/an
amount/quantity of*

Non-count quantificational nouns selecting a plural oblique

*dozens, scores, tens, hundreds, thousands, millions,
billions, zillions*

Even had Riordan or others implemented these distinctions in their coding, however, Huddleston and Pullum (henceforth H&P) still group together items that are further separated under the present study’s *Weak Number Hypothesis*, based on determiner singular/plural compatibilities. For example, H&P consider both *many* and *much* to be “degree determinatives,” while *a few* and *a little* are both “positive paucal determinatives” (H&P, 2002). The current work’s schema alternatively recognizes the commonality of plural-compatibles *many* and *a few*, while separately grouping *much* and *a little* in a singular-compatible indefinite quantifier category. Table 5 presents the codes used in the present study’s proposed determiner-type factor, along with examples drawn

from the study sample.

Table 5. Determiner-type factor coding.

Definite article (DEF)	
There's <u>the</u> boys right there.	(STP545JU091:S3)
Possessive (POSS)	
There's like <u>his</u> ears ...	(LES320SU085:S9)
Cardinal number (NUM)	
There's <u>two</u> pivotal concepts that I'm going to examine today.	(STP450SG128:S3)
Demonstrative (DEM)	
There are <u>these</u> sociological problems ...	(COL425MX075:S2)
No (NO)	
There're <u>no</u> collisions or anything of that sort.	(LES485MG006:S1)
Plural-compatible indefinite quantifier (PL-QNT)	
There're so <u>many</u> fun things to do.	(LAB175SU026:SU-f)
Number-transparent indefinite quantifier (Neut-QNT)	
There's <u>a lot of</u> things I don't mean to say out loud.	(LEL195SU120:S1)
Singular-compatible indefinite quantifier (SG-QNT)	
There's so <u>much</u> auction literature ... ⁴	(MTG270SG049:S1)
Indefinite article (SG-Indef)	
There is <u>an</u> exchange of different cultures ... ³	(MTG999ST015:S4)
Bare (no determiner)	
There's people from all over the world ...	(DIS115JU087:S1)
Other (including adjectives)	
There's <u>varying</u> degrees of autonomy ...	(STP450SG128:S3)

⁴ Because *SG-QNT* and *SG-Indef* require singular NPs, the examples given for these types are not actually drawn from our 500-token sample. Recall that *there+be* with singular NP was excluded, and as it happens, there simply were no tokens of these types with a plural NP. It seems likely that speakers find such a combination—singular determiner with plural NP—infelicitous quite independent of copula form.

Cue Distance

Once again, we define the novel factor Cue Distance (CD) as the distance in words from copula to the first temporal element that provides an unambiguous signal with regard to the plurality of the associated post-copular NP. The sentences in (4) provided some initial examples of calculating CD. There are still several types of determiners—those that provide no particular number signaling—where despite the presence of such an intervening element, Cue Distance will still equal the full distance to the post-copular head noun. In these cases, the CD ends up equal to the “Noun Distance” already factored into most prior studies. These types include both some “weak” and some “strong” forms from the Milsark (1977) taxonomy, including: definite articles, possessives, *no*, and number-transparent indefinite quantifiers (e.g., *some*), in addition to the “bare” and “other (including adjectives)” categories.

At the same time, there are four categories where the determiner itself provides an unambiguous signal, or cue, as to the grammatical number of the upcoming head noun. In these cases—including demonstratives, both plural-only and singular-only (mass noun) quantifiers, and indefinite articles—Cue Distance will now be just the distance to the determiner itself, regardless of the further distance to the head noun.

Finally, there is the curious case of cardinal numbers. Any number other than one might seem to provide an unambiguously plural signal, but consider the sentences in (20):

- (20) a. There are | *is twenty-two more students arriving tomorrow.
 b. There *are | is twenty-two percent more information in the next release.

If speakers' awareness of the possibility of cases such as (20.b) is part of their grammatical competence, Cue Distance for cardinal numbers would seem to measure neither to the determiner nor to the head noun, but rather to a point in between, that is, the underlined words in (20). In terms of the *Weak Number Hypothesis*, this interpretation would characterize a cardinal number as being a somewhat weaker plurality cue than an unambiguously plural element such as quantifier *several* or demonstrative *these*, yet stronger than a numerically blanded quantifier such as *some* or definite article *the*. This does not seem particularly convincing intuitively, but it would account for the results of prior studies in which numeric determiners have been found to slightly favor discord. In that light, cardinal number Cue Distance is coded in the present study as suggested by the underlining in (6) to see if our current data may support this approach.

To maintain a reasonable balance of tokens in each category, CD was encoded with values of “1,” “2,” “3-4,” or “5+.”

Other Measures

As in several prior studies, a measure of the distance from copula to post-copular head noun (Noun Distance, ND) was also still encoded—that is, in addition to Cue Distance—using the same range of values as CD. Determiner Distance (DD) was also encoded along similar lines for all tokens having a determiner—meaning all except the “bare” and “other” categories.

Lastly, a Cue Delta (CΔ) measure was encoded to capture the degree to which the presence of a given determiner type tends to extend CD. This corresponds to how much further out beyond the determiner a speaker must plan, depending on the type of

determiner. $C\Delta$ is defined as Cue Distance less Determiner Distance, as in (21).

$$(21) \quad C\Delta = CD - DD$$

By definition, $C\Delta$ ranges from zero, in cases where the determiner itself provides an unambiguous plurality cue (see example 22.a), to a maximum value of one less than Noun Distance, in cases where the determiner immediately follows the copula, but provides no incremental plurality information (22.b and 22.c). $C\Delta$ is undefined for tokens with no determiner. Table 6 summarizes the factors coded for this portion of the study, while (22) offers some examples of the calculation of ND, DD, CD, and $C\Delta$.

Table 6. Factors coded for analysis of determiner-type and “cue distance” effects.

<u>Factor</u>	<u>Abbreviation</u>	<u>Description</u>
Noun Distance	ND	= # of words from copula to head of the post-copular NP
Determiner Distance	DD	= # of words from copula to determiner in post-copular NP Ø if no determiner present (“bare” and “other” categories)
Cue Distance	CD	= # words from copula to first unambiguous signal as to NP number
Cue Delta	$C\Delta$	= $CD - DD$ Ø if DD is Ø

- (22) a. There are-copula many-det different equations._N that have been put forth ...
(DD = 1; ND = 3; CD = 1; CΔ = 0) (DEF305MX131:S2)
- b. There's s-copula the-det assumptions._N right in that blue box.
(DD = 1; ND = 2; CD = 2; CΔ = 1) (OFC575MU046:S1)
- c. There're re-copula some-det either social or technological constraints._N ...
(DD = 1; ND = 6; CD = 6; CΔ = 5) (MTG270SG049_S3)

Hypotheses

The WNH offered two specific predictions to be tested (3.i/ii): First, that greater CD will correlate with higher rates of discord; and second, that the rate of discord associated with a given determiner type will correlate with its average CΔ.

Results

Overall agreement. The average rate of discord among the 500 tokens in our MICASE sample of *there+be* existentials with plural post-copular NPs is 45.0% (225/500 tokens; see Table 7). While this figure is in line with the rates found by researchers working with other corpora in several of the earlier studies, and substantially lower than some—as high as 71% in Meechan and Foley (1994)—it is somewhat higher than the 39.7% (604/1520 tokens) that Riordan (2007) found in his coding of the full MICASE corpus, and the difference just crosses our $p = 0.05$ threshold of significance ($\chi^2 = 3.86$; $df = 1$; $p = 0.049$). This initially suggests that our 500 tokens might not comprise a fully representative sample. Another factor, though, could be that the present study excludes tokens from non-native speakers, whereas Riordan does not mention doing so in creating his database. Further, it appears that the MICASE corpus grew by about 150,000 words, a 9% increase, between when it was accessed for Riordan's study and for

the present one, which could also account for some portion of the variance.

Table 7. Rates of discord by existential form (with plural NPs).

Existential form	Discord	Concord	Total
<i>there's</i>	222		222
<i>there is</i>	3		3
<i>there're</i>		80	80
<i>there are</i>		195	195
Total	225 45.0%	275 55.0%	500

Noun distance. Prior studies have reported mixed results with regard to the ND effect on non-agreement. All agree that greater ND is generally associated with higher rates of discord, suggesting that some degree of processing effect is in play—the assumption there being that the farther in advance the speaker has to plan at the time of copula production, the greater the difficulty involved in calculating agreement. However, Meechan and Foley (1994), Hay and Schreier (2004), and most recently Riordan (2007)—in his study also using MICASE—find that while correlated, the effect does not actually rise to the level of statistical significance. Not surprisingly then, the current study similarly finds a correlation between increased ND and higher rates of discord (see Figure 1), but the effect remains statistically insignificant ($\chi^2 = 4.52$; $df = 1$; $p = 0.21$), despite the clearly visible general association.

Figure 1. Rates of discord for Cue Distance and Noun Distance against sample average (N = 500).

Cue Distance. As Figure 1 likewise illustrates, increased CD is, like ND, also associated with higher rates of discord, and as predicted, the CD effect is more pronounced than for ND. Moreover, while the difference between the CD and ND effects—the separation between the curves—may not be visually arresting in the graphic, the CD effect is statistically significant ($\chi^2 = 9.49$; $df = 1$; $p = 0.023$), as predicted by the WNH (3.i), where the ND effect was not significant. This suggests, once again, that intervening elements that provide strong number cues tend to promote concord.

Determiner Type. Figure 2 illustrates percentages of non-agreement by type of determiner, which appear to indicate that determiner type and agreement are in fact related. No tokens within the sample represent the strongly singular forms from the determiner taxonomy: SG-QNT (e.g., *much*) or SG-Indef (*a/an*). There were also very few (less than 10) demonstratives, definite articles, or possessives, and for this reason,

these types are all excluded from tests of statistical significance. With this noted, it is immediately clear that discord percentages vary significantly by determiner type ($\chi^2 = 19.32$; $df = 3$; $p = 0.0002$).

Figure 2. Percentage of non-agreement by type of determiner (asterisk indicates $0 < N < 10$ tokens).

Determiner Type and Cue Distance. Returning to the *Weak Number Hypothesis*, its second prediction (3.ii) is that—in terms of the factors coded for the present study—the rate of discord associated with a given determiner type will correlate with its average Cue Delta; in other words, the farther out beyond the determiner itself that a given type tends to require the speaker to plan (i.e., the $C\Delta$ measure), the greater the chance of discord. Strikingly, this is precisely what the results illustrated in Figure 3 appear to confirm.

Figure 3. As predicted by the WNH, determiner-type relative discord percentages align closely with associated Cue Deltas (asterisk indicates $0 < N < 10$ tokens).

With the determiner types arrayed in order by their respective percentages of discord—here presented as a positive or negative relative to the average discord found for the full sample (represented by the 45% horizontal midline on the graph)—the corresponding Cue Delta ($C\Delta$) values are also nearly perfectly aligned.⁵ In fact, the only types where $C\Delta$ and discord percentage do not align are POSS and DEF, which have just one and nine tokens, respectively, within the sample. As the dashed vertical midline indicates, determiner types with average $C\Delta$ values greater than 1.00—that is, those that bear a degree of number ambiguity that, on average, requires the speaker to plan more than one word further out into the post-copular NP beyond the determiner—are

⁵ As previously noted, limiting ourselves to plural post-copular NPs drives the absence of singular-compatible indefinite quantifiers (SG-QNT) and indefinite articles (SG-Indef) in our sample. The SG-QNT and SG-Indef columns ($N = 0$) in Figure 3 should be understood in this context.

associated with greater than average discord percentages, while those with $C\Delta$ values less than 1.00 are associated with lower than average percentages of discord.

The results of Study 1 appear to provide substantial support for the *Weak Number Hypothesis* in English existential post-copular agreement contexts. We may as yet gain further confidence in the robustness of these effects if we can confirm their presence in other commonly used corpora. The next study undertakes this challenge.

Chapter Four – Study 2: Multi-Corpus Investigation

For corroboration of the results of Study 1, a further study was carried out using three other American English spoken corpora: Switchboard, Callhome, and Charlotte. Together with MICASE, these bodies comprise the entire spoken-discourse contents of the American National Corpus (ANC), Second Release (Reppen et al., 2005).

Switchboard

The version of the corpus used in the present study includes roughly 2.9 million words drawn from a collection of about 2,400 phone conversations recorded by Texas Instruments in 1990. Randomly connected pairs of subjects were asked to discuss a given topic prompt (Reppen et al., 2005).

As with MICASE in Study 1, a random sample was extracted, using sentences with existential uses of *there is*, *there's*, *there are*, and *there're* followed by an associated plural NP. In this case, a total of 507 tokens were processed, using the same exclusions, tokenization procedures, classification and encoding schemes, including head-noun distance (ND), determiner distance (DD), Cue Distance (CD), and Cue Delta (CΔ).

Overall Agreement. The average rate of discord in the Switchboard sample is 63.1% (320/527), considerably higher than the 45.0% observed in the MICASE sample.

Noun and Cue Distances. Once again as in the MICASE study, both ND and CD exhibit a general association with discord in the Switchboard data (see Figure 4), with the CD effect, as predicted, being somewhat more pronounced. The ND correlation remains statistically insignificant ($\chi^2 = 5.58$; $df = 1$; $p = 0.13$), but here the CD association

also falls just short of statistical significance ($\chi^2 = 7.71$; $df = 1$; $p = 0.052$).

Figure 4. Rates of discord for Cue Distance and Noun Distance in the Switchboard sample ($N = 507$).

Determiner Type. In the Switchboard sample we again find no singular-only indefinite quantifiers (SG-QNT) and very few demonstratives, definite articles, indefinite articles, or possessives, all of which are once again excluded from tests of significance. Whereas the overall variance by determiner type was quite significant in the MICASE sample, in the Switchboard data (see Figure 5) it falls slightly short of significance ($\chi^2 = 5.23$; $df = 2$; $p = 0.073$). What seems most significant here, however, is that the plural-only indefinite quantifiers (PL-QNT) and number-transparent indefinite quantifiers (Neut-QNT), while not as dramatically distinct as in the MICASE study do continue to reflect statistically different influences on discord ($\chi^2 = 8.06$; $df = 1$; $p = 0.0045$). It is these categories, along with singular-only indefinites, that have most often been kept together in the determiner taxonomies of prior studies.

Figure 5. Non-agreement by type of determiner, Switchboard sample (“*” = $0 < N < 10$ tokens; “**” = 0 tokens).

Determiner Type and Cue Distance. Here we are looking again to see if rates of discord by determiner type correlate with how far beyond the determiner the corresponding cue distances lie on average (i.e., the average $C\Delta$ metric for a determiner type should correlate with its discord percentage). After excluding the categories with fewer than ten total tokens in the sample (SG-QNT, DEM, DEF, and SG-Indef), the remaining categories (see the enclosing rectangle in Figure 6) exhibit the expected alignment. As in Figure 3 for MICASE, the dashed vertical line shows that Switchboard determiner types with average $C\Delta$ values greater than 1.00—requiring the speaker to plan at least one word beyond the determiner for number agreement—are associated with greater than average discord percentages, while those with $C\Delta$ values less than 1.00 are associated with lower than average percentages of discord.

Figure 6. Cue Delta alignment with determiner-type discord percentages for Switchboard sample.

Charlotte and Callhome Corpora

The final two spoken components of the American National Corpus are comparatively small. The Charlotte Narrative and Conversation Collection (Charlotte) forms a collection of 93 narratives, conversations, and interviews—a total of about 190,000 words—collected about the vicinity of Charlotte, North Carolina. The portion of the Callhome corpus available in the ANC presents just a little less than 50,000 words in 24 telephone conversations between friends and family. While small, each corpus offers a slightly different style of discourse and can thus help in testing the robustness of the effects found in Study 1.

Whereas with MICASE and Switchboard, random samples were extracted, for Charlotte and Callhome, the smaller set of tokens available was coded in its entirety. Once again the same method, encoding scheme, and exclusions were employed. The Charlotte corpus yielded a total 72 examples of existential *there+be* constructions having

a plural associated post-copular NP, while Callhome produced just 28 such tokens.

Overall Agreement. The average rates of discord were 56.9% among the Charlotte data and 60.7% for Callhome.

Determiner Type and Cue Distance. When separated by determiner type, most categories in both corpora held too few tokens for valid tests of statistical significance, but a final four-corpus comparison of the relationship between the Cue Delta metric and rates of discord by determiner type is illustrative of the general trends. Figure 7 charts percentage rates of non-agreement relative to the average discord rate for each corpus. Categories are excluded individually by corpus if they hold fewer than ten tokens in MICASE or Switchboard (where total sample size equals 500 and 507 tokens, respectively) or fewer than five in Charlotte or Callhome. Only MICASE exceeded this threshold for all six determiner types; Callhome has just two types that meet the five-token threshold.

Figure 7. Comparative discord rates and corresponding average Cue Delta values for the four corpora (excludes types with $N < 10$ for MICASE and Switchboard, $N < 5$ for Charlotte and Callhome).

Discussion

We can make a few observations from these data. First, it was hypothesized that demonstratives and indefinite quantifiers that are compatible only with plural NP associates (type PL-QNT, e.g., *many*, *a number of*)—strong in terms of their number semantics—provide an unambiguous signal as to the ultimate plurality of the upcoming NP head. Correspondingly, these types have the lowest average Cue Delta scores in Figure 7 for all types across all corpora (in which they are represented), and the WNH predicts that these elements, when present, should promote concord. The results confirm

that these categories are, indeed, associated with the lowest rates of discord across all corpora.

At the same time, it was suggested herein that negation particle *no*, definite article *the*, and indefinite quantifiers that are compatible with either plural or singular NP associates (type Neut-QNT, e.g., *some*, *a lot*) are semantically blanched in terms of number signaling, providing the least information with regard to the plurality of the associated NP head. Correspondingly, these types have the highest average CΔ scores, and the WNH predicts that when present, they will tend to promote discord. Once again, Figure 7 reveals that collectively these categories are associated with the highest rates of discord across all corpora.

The sole exception to this is in Switchboard, where the *No* category dips below cardinal numbers (NUM) in terms of discord rate. This correctly corresponds, however, to the fact that the average CΔ for *No* within Switchboard is also lower than for NUM. Speculating, Switchboard, as represented by the 507-token sample, could have a higher proportion of weighty compound numbers (e.g., *eighty-seven thousand*), which would tend to increase average CΔ. Alternatively, the sample might have comparatively short average distances between *no* and head noun, lowering average CΔ for type *No*. Such variation in order among these middle categories aside, though, even Switchboard displays greater average CΔ and correspondingly higher rates of discord for number-transparent indefinite quantifiers (Neut-QNT) than for plural-only quantifiers.

Collectively, these results provide evidence that the effects observed in Study 1 are robust across a range of spoken American English corpora and appear to provide

strong support for the present work's central hypothesis.

Chapter Five – Study 3: Modeling the Effects on Discord for New Factors

Multiple logistic regression analysis in syntactic studies (Chapter 2) offers a method for modeling variation in production by simultaneously considering a group of factors that are each known to be separately predictive. The goal is to discern the relative strength of each factor while statistically controlling for the effects of the others.

In Riordan 2007, in particular, the author produces such a model of agreement variation in English existential *there+be* constructions, using logistic regression and the MICASE corpus. Riordan develops both a fixed-effect model and a mixed model of the phenomenon. The purpose of the latter is to better generalize from the choices of a specific set of speakers to an estimated behavior for the population at large.

The *Weak Number Hypothesis* is primarily concerned with providing an explanatory view of the strong effect of determiner type observed (but largely unexplained) in prior studies of *there+be* agreement variation. WNH does, however, assert that one novel factor, Cue Distance, should be predictive of discord. While Studies 1 and 2 provide evidence of this predictive capacity, it also follows from WNH that the addition of CD to an existing regression model should improve its overall fit and predictive accuracy.

Model and Coding

As an initial step before integrating our new factors, the existing model from Riordan (2007) is reproduced using our reduced sample from MICASE ($N = 500$). We begin with Riordan's final, most predictive model: the mixed model that incorporates a

random variable to control for the baseline agreement bias of each individual speaker. The factors that Riordan coded were listed in Table 1. Of these, his reduced mixed model included five as having significant effects: age, primary discourse mode, weight of post-NP sequence, and certain combinations of determiner type and distance between the copula and the head of the post-copular NP.

Age and discourse mode are already encoded in the MICASE source data. Taking up once again the 500-token sample from Study 1, the remaining factors were annotated using a combination of custom software and hand coding. To replicate the existing model, Riordan's taxonomy for determiner type was used initially rather than the present work's alternative scheme proposed above in Study 1. Riordan's encoding of distances was also preserved, wherein he used a four-level factor for some elements and a reduced, binary classification for others ("Minimum" for a distance of a single word or "Extended" for distances of two or more).⁶ One inevitable and potentially significant change comes in that, as Riordan mentions in his endnotes, a new finer-grained recoding of the discourse mode factor was introduced to the online version of MICASE just as Riordan 2007 was going to press. This change was not reflected in Riordan's data, but it is unavoidably part of the sample on which our replication of his model is based.

Table 8 lists the resulting fixed-factor coefficients after fitting a Generalized Linear Mixed Model (Baayen, 2007; Baayen, Davidson, & Bates, 2008), with *p*-values reported for those elements that reached a 95% confidence threshold for significance.⁷

⁶ See Riordan (2007) for a detailed discussion of the author's factor levels and encoding.

⁷ Regression models were developed using R version 2.8.1 (R Development Core Team, 2006).

Table 8. Reproduction of Riordan (2007) mixed-effects model fitted to Study 1 sample ($N = 500$).

	Estimated Coefficient (β)	p -value
(Baseline)	0.23	
Age = 24-30	-0.28	
Age = 31-50	0.27	
Age = 51+	1.21	< 0.01
Age = unknown	-15.11	
Discourse mode = Highly monologic	1.35	< 0.05
Discourse mode = Moderately interactive	0.30	
Discourse mode = Moderately monologic	0.91	< 0.01
Discourse mode = Mixed	0.59	
Post-NP sequence = Minimum	-0.41	< 0.05
Det type/dis = Definite/Extended	-0.58	< 0.05
Det type/dis = Definite/Minimum	0.15	
Det type/dis = Indefinite/Extended	-0.42	< 0.01
Det type/dis = Indefinite/Minimum	-0.43	
Det type/dis = NCQN/Extended	-1.07	< 0.05
Det type/dis = NCQN/Minimum	-1.10	
Det type/dis = <i>No</i> /Extended	-1.46	
Det type/dis = <i>No</i> /Minimum	-0.27	
Det type/dis = None/Extended	-0.04	< 0.01
Det type/dis = None/Minimum	-0.85	
Det type/dis = Num/Extended	-0.95	
Det type/dis = Num/Minimum	-0.44	

To gauge how well we have replicated Riordan's model, we can use the same measure that he cites: classification accuracy. As shown in Table 9, the ability of the replicated model to predict concord vs. discord stands at 74.8%, which is somewhat lower than the 80.1% figure reported by Riordan (2007).

Table 9. Classification accuracy on Study 1 sample ($N = 500$).

		<u>Predicted</u>		
		Discord	Concord	Correct
<u>Observed</u>	Discord	150	75	66.7%
	Concord	51	224	81.5%
Overall:				74.8%

There are several possible explanations for these differences. The first was mentioned above: the change in MICASE's encoding of discourse mode. Style of discourse was one of the most significant factors in the Riordan 2007 model, and any change in its coding could certainly have a substantial effect. Next, as discussed in Study 1, tokens from non-native speakers were excluded, whereas Riordan does not mention having done so. Another obvious possibility is that our 500-token sample may not represent the same distribution of factors as Riordan's exhaustive coding of the full corpus ($N = 1520$). If so, a different set of factors fed to the newer regression might yet produce a better fit. It is certainly also possible that Riordan's encoding scheme has been inadvertently implemented in a slightly different manner here.

These speculations aside, it should be noted that the aim of the present study is neither simply to identically reproduce the results of a particular prior study nor even to generate the best possible predictive model for *there+be* agreement; rather the goal in replicating Riordan (2007) is to see if a good model can be made better through the addition of either Cue Distance or the new plurality-oriented determiner taxonomy or both. The replicated model thus suffices for the present purpose.

Modeling Cue Distance

With this reproduction of the Riordan 2007 model in hand, the next step is to gauge the effect of adding Cue Distance to the regression. As we saw in Study 1, CD is significantly correlated (or “collinear”) with Noun Distance (see Figure 1), a component of the existing model. Because such collinearity is problematic for regression models (Baayen, 2007; Jaeger, 2007), we remove Riordan’s combined Determiner Type/Distance factor and replace it with CD, rather than just adding the new measure. Table 10 outlines the updated model.

Table 10. Mixed-effects model updated to include Cue Distance factor.

	Estimated Coefficient (β)	<i>p</i> -value
(Baseline)	0.74	< 0.05
Age = 24-30	-0.18	
Age = 31-50	0.25	
Age = 51+	1.27	< 0.01
Age = unknown	-15.25	
Discourse mode = Highly monologic	1.56	< 0.05
Discourse mode = Moderately interactive	0.37	
Discourse mode = Moderately monologic	0.98	< 0.05
Discourse mode = Mixed	0.71	
Post-NP sequence = Minimum	-0.33	
Cue Distance = 2	-1.12	< 0.001
Cue Distance = 3-4	-1.82	< 0.0001
Cue Distance = 5+	-1.96	< 0.0001

As the coefficients and *p*-values indicate, CD now yields the largest-magnitude and most statistically significant factors in the updated model. As seen in Table 11,

classification accuracy achieves a modest gain, rising to 76.6% (+1.8%).

Table 11. Classification accuracy with Cue Distance added to regression.

		<i>Predicted</i>		
		Discord	Concord	Correct
<i>Observed</i>	Discord	163	62	72.4%
	Concord	55	220	80.0%
Overall:				76.6%

While the improvement in model fit is not dramatic on its own, the fact that this increase comes with noun distance and determiner type entirely removed from the model supports once again the initial prediction of the WNH (3.i): that CD will not only correlate with discord, but specifically, it will have a more significant effect than does overall distance to the head noun.

Modeling Determiner Type

As a final exercise, determiner type is added back into the model, now based on the plurality-oriented taxonomy outlined in Study 1. Because determiner type and Cue Distance are significantly correlated (i.e., collinear), we can follow the approach taken by Riordan (2007) in creating a single, combined factor: Determiner Type / Distance. This final model produces exactly the same classification accuracy (76.6%) as the prior regression without a measure of Determiner Type.

Chapter Six – Discussion

The apparent correlation between determiner type and non-agreement in English *there+be* existentials has been noted by nearly every prior variationist study on this topic (Britain & Sudbury, 2002; Hay & Schreier, 2004; Meechan & Foley, 1994; Riordan, 2007; Tagliamonte, 1998; inter alia). Riordan goes so far as to suggest a cause-and-effect relationship: “...speakers are aware of the collocational statistics of *there’s* ... Thus, concord patterns may be determined by the shape of NPs because these statistics provide biases in online processing” (Riordan, 2007:272). Study 3 showed, however, that determiner type holds no benefit for the mixed-effect regression model, once Cue Distance has been added, which suggests that the correlation between type of determiner and discord is at best indirect. Determiner type may affect Cue Distance, but ultimately, it appears that it is the extended Cue Distance itself that promotes discord, regardless of which type of determiner, if any at all, is involved.

To be clear, Study 1 did show that a determiner, if present, can affect Cue Distance, at least in one direction; as suggested by the *Weak Number Hypothesis*, determiners that are strong in terms of their number-signaling semantics can substantially shorten Cue Distance by providing an advance signal as to the plural or singular nature of the post-copular head noun. This effect was illustrated by the introductory examples in (4). With distance to head noun held constant, the replacement of semantically blanded indefinite quantifier *some*—sitting immediately adjacent to the copula—by strongly plural yet equally indefinite determiner *several*, cuts total Cue Distance to be equal to just Determiner Distance (DD = 1 in these examples), down from Noun Distance (ND = 5 in

these examples).

In the other direction, certainly the insertion of non-plurality-signaling words, whatever they may be, between copula and head noun increases total Cue Distance. If other elements are held constant, the simple increase in number of words (or “weight” as it is frequently termed) alone does correlate with discord. Again, though, such is the case for any category, not an effect peculiar to determiners. This does tend to undercut one aspect of the *Weak Number Hypothesis*, suggesting that while plurality-signaling determiners do promote concord by shortening Cue Distance, number-transparent determiners, on the other hand, only promote discord to the extent that the inclusion of the words themselves increases the total weight of the phrase, not as a reflection of speaker knowledge of collocation statistics.

These findings would seem to lend support to the suggestion noted above from Cheshire (1999) that having a limited capacity for “look ahead,” speakers are frequently led by discourse pressures to select singular agreement—particularly in the form of contracted *there’s*—as the more basic option. Cheshire argues that prescriptively driven plural agreement only takes hold when speakers have sufficient time—for example in monologic discourse—to plan farther ahead into the post-copular NP. At the same time, the fact that stronger semantically plural determiners appear to actively promote concord by reducing Cue Distance seems to argue against Cheshire’s suggestion (also Crawford, 2005) that speakers may not actually compute agreement when constructing *there+be* expressions. On that view, it is suggested that speakers might be selecting one or another (*there is/’s* vs. *there are/’re*) formulaic, lexicalized construction that involves collocation

patterns with certain determiners. The results of Study 3—where the inclusion of Determiner Type makes no improvement over the inclusion of Cue Distance alone—would tend to argue against such a position.

This finding also places English *there+be* non-agreement slightly out of line with certain other variation phenomena. Work with the English dative shift (Bresnan, 2007), relativizer and complementizer optionality (Jaeger, 2006; Wasow, Jaeger, & Orr, 2011), and other alternations offers evidence that speakers may make choices among syntactic alternatives at least in part based on knowledge of collocation statistics included as part of their grammatical competence. Agreement variation may be an exception to this trend. While nothing in the current study argues *against* such a view of competence, neither does the present work offer any new evidence in support of a suggestion “for including some quantitative information in the grammar of English” (Wasow, 2009).

At the same time, the conclusion that greater Cue Distance promotes discord is certainly in line with Hawkins’ principle of *Minimize Domains* (Hawkins, 2001, 2004), which argues that the human processor generally prefers to keep to a minimum the spans between semantically or syntactically linked components. The significance of the CD effect on agreement, as shown in the present work, suggests that in the post-copular context of English existential *there+be* constructions, not only is there such a linkage between copula and post-copular head noun, but that in terms of supporting agreement, there is an even more critical linkage between the copula and the first post-copular number-signaling element. This also places the current study in line with other multivariate syntactic analyses that offer processing-based explanations (Gries, 2003;

Lohse, Hawkins, & Wasow, 2004).

Chapter Seven – Conclusion

Prior work by a number of researchers has offered substantial evidence to suggest that agreement variation with English existential *there+be* is subject to quite a range of factors, including linguistic, processing, social, and discourse elements. The present study proposes a novel explanation for the significance of one such factor: a correlation between type of determiner and non-agreement. As motivated by a *Weak Number Hypothesis*, two additional factors have been proposed: the number of words from copula to first unambiguous number signal (Cue Distance); and an original taxonomy for type of determiner based on plurality semantics rather than an alignment by definiteness. Distributional analyses across four different spoken corpora and a mixed-effects logistic regression analysis have provided evidence of the significance of these factors. Most interesting, perhaps, is the conclusion that the observed correlation of discord and determiner type may be subsumed under a processing explanation based solely on the effect of Cue Distance, with determiner type now posited as having only an indirect role.

While these investigations further the discussion of the phenomenon, there are certainly still any number of factors that remain for further study. In particular, there is the body of literature on issues affecting agreement in non-existential English declarative sentences, including as noted Chapter 2: structural priming (Bock, 1986; Bock & Loebell, 1990), variant concord with collectives (Bock, Nicol, & Cutting, 1999; den Dikken, 2001; Humphries & Bock, 2005), and the effect of pluralia tantum (Bock et al., 2001), among others. It would be informative to factor these elements *out* of models for *there+be* non-agreement as these might be considered to be more general effects not

peculiar to the alternation explored in the present work.

Finally, with comprehensive multivariate models for *there+be* agreement variation available (Riordan, 2007, as augmented by the present work), this would seem to be an area ripe for psycholinguistic experimentation, with the aim being to provide further corroboration of the several effects suggested by this and prior research. Past, present, and potential future work can further our understanding of what falls within competence, what is attributable to processing, and where the line between them—if there is one—lies.

References

- Baayen, R. H. (2007). *Analyzing linguistic data: A practical introduction to statistics*. Cambridge: Cambridge University Press.
- Baayen, R. H., Davidson, D. J., & Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory & Language*, 59(4), 390–412.
- Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, 18(3), 355–387.
- Bock, K., & Miller, C.A. (1991). Broken agreement. *Cognitive Psychology*, 23, 45–93.
- Bock, K., & Loebell, H. (1990). Framing sentences. *Cognition*, 35(1), 1–39.
- Bock, K., Nicol, J., & Cutting, J. C. (1999). The ties that bind: Creating number agreement in speech. *Journal of Memory and Language*, 40(3), 330–346.
- Bock, K., Eberhard, K. M., Cutting, J. C., Meyer, A. S., & Schriefers, H. (2001). Some attractions of verb agreement. *Cognitive Psychology*, 48, 83–128.
- Breivik, L. E. (1990). *Existential there: A synchronic and diachronic study*. Oslo: Novus Press.
- Breivik, L. E., & Martinez-Insua, A. E. (2008). Grammaticalization, subjectification and non-concord in English existential sentences. *English Studies*, 89(3), 351–362.
- Bresnan, J. (2007). Is syntactic knowledge probabilistic? Experiments with the English dative alternation. In S. Featherston & W. Sternefeld (Eds.), *Linguistics in search of its evidential base* (pp. 75–96). Berlin: Mouton de Gruyter.
- Bresnan, J., Cueni, A., Nikitina, T., & Baayen, R. H. (2007). Predicting the dative alternation. In G. Boume, I. Kraemer, & J. Zwarts (Eds.), *Cognitive Foundations of Interpretation* (pp. 69–94). Amsterdam: Royal Netherlands Academy of Science.
- Britain, D. (2002). Diffusion, levelling, simplification and reallocation in past tense BE in the English Fens. *Journal of Sociolinguistics*, 6(1), 16–43.
- Britain, D., & Sudbury, A. (2002). There's sheep and there's penguins: 'Drift', 'slant' and singular verb forms following existentials in New Zealand and Falkland Island English. In M. C. Jones & E. Esch (Eds.), *Language change: The interplay of internal, external, and extra-linguistic factors* (pp. 211–240). New York: Mouton de Gruyter.

- Cheshire, J. (1999). Spoken standard English. In T. Bex & R. J. Watts (Eds.), *Standard English: The widening debate* (pp. 129–148). New York: Routledge.
- Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.
- Crawford, W. (2005). Verb agreement and disagreement: A corpus investigation of concord variation in there + be constructions. *Journal of English Linguistics*, 33(1), 35–61.
- Dikken, M. den. (2001). ‘Plurilinguals’, pronouns and quirky agreement. *The Linguistic Review*, 18, 19–41.
- English Language Institute. (2007). *MICASE manual: The Michigan corpus of academic spoken English, version 3, work in progress*. R. C. Simpson, D. Y. W. Lee & S. Leicher, revised by A. Ädel. The University of Michigan.
- Godfrey J., Holliman, E., & McDaniel, J. (1992). SWITCHBOARD: Telephone speech corpus for research and development. *Proceedings of ICASSP-92*, San Francisco, 517–520.
- Gries, S. Th. (2003). *Multifactorial analysis in corpus linguistics: A study of particle placement*. London, New York: Continuum Press.
- Hawkins, J. A. (2001). Why are categories adjacent? *Journal of Linguistics*, 37(1), 1–34.
- Hawkins, J. A. (2004). *Efficiency and complexity in grammars*. Oxford: Oxford University Press.
- Hay, J., & Schreier, D. (2004). Reversing the trajectory of language change: Subject–verb agreement with be in New Zealand English. *Language Variation and Change*, 16, 209–235.
- Huddleston, R., & Pullum, G. K. (2002). *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Humphreys, K. R., & Bock, K. (2005). Notional number agreement in English. *Psychonomic Bulletin & Review*, 12(4), 689–695.
- Jaeger, T. F. (2006). *Redundancy and syntactic reduction in spontaneous speech*. Dissertation, Stanford University.
- Kingsbury, P., Strassel, S., McLemore, C., & MacIntyre, R. (1997). *CALLHOME American English lexicon (PRONLEX)*. Philadelphia: Linguistic Data Consortium.

- Labov, W. (1964). Phonological correlates of social stratification. *American Anthropologist*, 66(6_PART2), 164-176.
- Lohse, B., Hawkins, J. A., & Wasow, T. (2004). Domain minimization in English verb-particle constructions. *Language*, 80(2), 238–261.
- Martinez Insua, A. E., & Palacios Martinez, I. M. (2003). A corpus-based approach to non-concord in present day English existential there-constructions. *English Studies*, 84(3), 262–283.
- Meechan, M., & Foley, M. (1994). On resolving disagreement: Linguistic theory and variation—There's bridges. *Language Variation and Change*, 6(1), 63–85.
- Milsark, G. (1977). Towards an explanation of certain peculiarities of the existential construction in English. *Linguistic Analysis*, 3, 1–31.
- Pietsch, L. (2005). *Variable grammars: Verbal agreement in northern dialects of English*. Tübingen: Niemeyer.
- Quirk, R., & Wrenn, C. L. (1957). *An Old English grammar, 2nd edition*. London: Methuen.
- R Development Core Team. (2004). *R: A language and environment for statistical computing*. <http://www.R-project.org>.
- Reppen, R., Ide, N., & Suderman, K. (2005). *American national corpus (ANC) second release*. Philadelphia: Linguistic Data Consortium.
- Riordan, B. (2007). There's two ways to say it: Modeling nonprestige there's. *Corpus Linguistics & Linguistic Theory*, 3(2), 233–279.
- Rupp, L. (2005). Constraints on non-standard –s in expletive there sentences. *English Language and Linguistics*, 9(2), 255–288.
- Sauerland, U., & Elbourne, P. (2002). Total reconstruction, PF movement, and derivational order. *Linguistic Inquiry*, 33(2), 283–319.
- Schilling-Estes, N., & Wolfram, W. (1994). Convergent explanation and alternative regularization patterns: Were/weren't leveling in a vernacular English variety. *Language Variation and Change*, 6(3), 273–302.
- Schütze, C. (1999). English expletive constructions are not infected. *Linguistic Inquiry*, 30(3), 467–484.

- Simpson, R. C., Briggs, S. L., Ovens, J., & Swales, J. M. (2002). *The Michigan corpus of academic spoken English (MICASE)*. Ann Arbor, MI: The Regents of the University of Michigan.
- Tagliamonte, S. (1998). Was/were variation across the generations: View from the city of York. *Language Variation and Change*, 10(2), 153–192.
- Wasow, T. (2009). Gradient data and gradient grammars. In *Proceedings of the 43rd Annual Meeting of the Chicago Linguistics Society*, 255–271.
- Wasow, T., Jaeger, T. F., & Orr, D. (2011). Lexical variation in relativizer frequency. In H. Simon & H. Wiese (Eds.), *Expecting the unexpected: Exceptions in grammar* (pp. 175–195). De Gruyter.

Appendix A – Corpus Sample

500-token sample drawn from MICASE, existential *there+be*

<u>#</u>	<u>Text</u>
1	there 's some things that i learned like really really well that way and there 's some things didn't.
2	there 's philosophers that do so that 's why they state it.
3	there 's yeah there 's tons of 'em.
4	there 're two due early papers.
5	there 's you know the various looks or whatever of these things.
6	there are a couple of equations that do a pretty good job at least at the low-temperature low.
7	there are people you know the the enforcement is in place and they have basically a sampling effort.
8	there are more young ones than each age class of old ones but they pile up in the size categories.
9	there 's fewer students to go around and so there 's almost a recruitment aspect.
10	there are two stars here that are bright enough that we should learn what they are.
11	there 's uh there 's two ways to do that right?
12	there 's some books back there by his dad you know.
13	there 're just a lot of uh factors.
14	there 's tons of 'em.
15	there 's two different things right.
16	there 's three thousand nanometers?
17	there are a number of places in the book where her omnipotence of thought is demonstrated.

- 18 there 're two shuttles what are the shuttles?
- 19 there 's pictures with less less clothing on.
- 20 there 's so many different cultures and groups that live there.
- 21 there are some new people saying hey wait i know something to that that can't be a law.
- 22 there are all these other options i want to explore like creative writing and and such.
- 23 there are some single sex schools on this file.
- 24 there 're lots of sea mounts and ridges and fractures on it and stuff like that it.
- 25 there 's tons of different levels of interpretation which are gonna happen.
- 26 there 're prep questions for that then you may w- wanna just jot this down on your syllabus pages.
- 27 there 's no water spots on those leaves.
- 28 there 're like four that just (xx) by the that bird box.
- 29 there 's there 're many differences.
- 30 there 's some rules of thumb about the size of the well that will affect that it depends upon transmission.
- 31 there 're two mechanisms to stop transcription.
- 32 there 's few there 's a few um genes that 're on Y like genes that specifically only males have.
- 33 there 's at least five of you who did that search here and i know you 're all very reluctant to speak.
- 34 there 's two things okay one where it 's always bound i- if it 's like it 's always not bound.
- 35 there are a lot of secret societies that have their meetings up there.
- 36 there 's so much like there are n't just a couple avenues to get information.
- 37 there 's a lot of things i don't mean to say out loud.

- 38 there are no other interactions on the surface that could make the roughness increase like that.
- 39 there 's all different of societal norms that you could see really explicitly in T-V programs.
- 40 there are a couple of other um of other things we may be able to do as well.
- 41 there are two types of slack-time intervals.
- 42 there 's different competing groups.
- 43 there 's the boys right there.
- 44 there 's fourteen other ones there.
- 45 there 're some fanatics in the class that will immediately want to do that.
- 46 there 's eight birds on the wire.
- 47 there 's all this free floating modalities.
- 48 there 're some either social or technological constraints.
- 49 there are presynaptic receptors.
- 50 there is a lot of different issues around individuals versus just participation.
- 51 there 's about ten different biology courses you could take.
- 52 there 's a couple words in Latin that are called postpositives.
- 53 there 're things that lie beyond your ability to to change.
- 54 there 's just so many receptor subtypes.
- 55 there 's a thousand things that can go wrong in an experiment.
- 56 there are a few planets that are really really bright right now.
- 57 there 's ethnicity based states.
- 58 there 's any universal morals.
- 59 there 's other ones too but these are the major ones that we 'll talk about.

- 60 there are so many authors that wrote about the camps.
- 61 there are very few of these germinal centers present as a matter of fact on this particular cross section.
- 62 there are one finger typists who go faster than that.
- 63 there 's all the different the schools lemme use for example the schools that are yeah.
- 64 there are a lot of homeless people and um.
- 65 there 's three of us and my homologue does too.
- 66 there are some events that are somewhat less probable than other events.
- 67 there are others of the Essays and the work in the History but there are also several obvious attacks.
- 68 there 's a few molecules around me now okay?
- 69 there are there are cases.
- 70 there 's parts of an experience you take the visual field one slice at a time.
- 71 there are dozens more i could give but let 's just use three or four as the you know representative sample.
- 72 there are two different views of something.
- 73 there 's none in those buckets.
- 74 there are other sigma factors um which are turned on in response.
- 75 there 's a whole bunch of questions you can ask.
- 76 there 's example exams on the web or on the web page for two eighty-one.
- 77 there 's similar components.
- 78 there 's the assumptions right in that blue box.
- 79 there 's lots of different things that you can ask questions about.
- 80 there 's so many different ways.

- 81 there 's maps here it 's in town so please take a map and um see us there about seven-thirty.
- 82 there are Basques in France yeah but not up in Paris where standard French was formed.
- 83 there are more of them in this room than there are of us.
- 84 there 's so many of them.
- 85 there 's like his ears and that 's his eye.
- 86 there 're rules about weather.
- 87 there 's three levels of data not two.
- 88 there 're always students saying no no i 'm not really staying that long i 'm staying longer.
- 89 there 's three doors.
- 90 there 're profound calcium changes that would actually be very useful uh to actually be able to.
- 91 there 's multiples of these um things.
- 92 there 's motor patrols all over the place it is a very very safe campus and i 'm very feel very safe.
- 93 there 's route tests that test somebody 's ability to memorize a route.
- 94 there are site conditions that during the day it 's not as hot because you 're adjacent to the water.
- 95 there are too many double negatives floating around.
- 96 there are concerns that it um infringes on um citizen rights when um uh endangered species occurred.
- 97 there are others such as how well does your knowledge uh.
- 98 there 's lots of children terms.
- 99 there are industrialists and workers in an industrial uh capitalist mode of production.

- 100 there are six teams i 'll do.
- 101 there 're two guys in relative motion so we 'll put one here in the ground.
- 102 there 're i can tell there 're some fanatics in the class that will immediately want to do that.
- 103 there 's so many like there 's so many things.
- 104 there 're a lot of different kinds of methods that are basically specialized.
- 105 there are times when you need more of some product there 's times when you need less of some product.
- 106 there are acceptable alphas iden- uh for each of the identified subscales.
- 107 there are the other facts.
- 108 there 're the second group of young women out there.
- 109 there 're categories concepts and rules.
- 110 there 's ninety-nine people left here that don't have the disease.
- 111 there 's six and six.
- 112 there 's two well the sentence right there there 's the two parts it 's valid and it 's invalid.
- 113 there are there are salons in Newton 's time.
- 114 there 's allusions to matrilineal and matrilocal sys- values.
- 115 there 's two levels of studio and they wanted to get light to the lower one.
- 116 there 's certain distortions that take place in the language.
- 117 there 's ten thousand of you anyway and there 's eight masses.
- 118 there are people also studying plants um invertebrates reptiles you name it.
- 119 there are the lounges.
- 120 there are social norms but then there 're also other factors that help people to choose.

- 121 there 's there 's uh there 's two ways to do that right?
- 122 there are choices that you could have made other ways.
- 123 there 're a bunch there- there 're many little things in here i 'd like you to just take a look at.
- 124 there are ten distracters it takes longer to find the T than if there 're only four distracters.
- 125 there 're different influences that can come in.
- 126 there 're two olfactory bulbs so there 's one on each side.
- 127 there 's sixteen it takes longer than ten and four so.
- 128 there are individuals who can use that time uh in prison to kick drugs.
- 129 there 's types of schemas.
- 130 there 're many things i can't do most things i can't do.
- 131 there are several of you who missed the comments on the paper.
- 132 there are times when people wanna use two different forms of the same test.
- 133 there 're just always people ready to spend the money on the things that they think are important.
- 134 there 're very few pieces that a museum would own where you have the backup sketches.
- 135 there 's still some things that i wanna put in that i just don't have time for.
- 136 there are no implications for the future.
- 137 there 's different types in pornography that are definitely racialized.
- 138 there 're constraints put on it.
- 139 there are state supported uh preschool programs as well but these are restricted to poor kids as well.
- 140 there are these reasons.
- 141 there 's a lot of different levels on which you can understand literature right?

- 142 there are less complicated things that 's an easier way to ask questions.
- 143 there are mental objects with nonmental features.
- 144 there 's lots of other serotonin receptors we 've found postsynaptically.
- 145 there are no rational foundations of the kind that Leibniz and Clarke in their different manners.
- 146 there 's that there 're more there 're other important issues to deal with.
- 147 there 's some parts in it where it says you can skip this
- 148 there are a- a lot of others that imply adult rather than child activity.
- 149 there are there are theories (oh) i don't know if we 'll get there this semester.
- 150 there are two big stone pumas in front of the Natural Science Museum.
- 151 there are many many many mechanisms by which that can that production can occur.
- 152 there are nucleophiles around but sometime- like if a nucleophile 's not strong enough to push out.
- 153 there 's a hundred balls or so and who knows how many possibilities.
- 154 there 's like forty fifteen-year-olds.
- 155 there are certain similarities between uh people of a different culture e- uh of a different group.
- 156 there 're many ways to test for a V-S-D uh cardiologists will often use a just a teth-stethoscope.
- 157 there 's a lot of stereotypes out there and i don't know if you know he 's just used to that.
- 158 there are um there are a couple of other um of other things we may be able to do as well.
- 159 there 's gender differences in reading for these little kids and girls are really doing a lot better.
- 160 there 's a few more types of validity that we 'll just sort of well there 's one more.

- 161 there are a lot of interesting issues uh that relate to what for example flight-management systems.
- 162 there 's a lot of bagels left so feel free.
- 163 there are certain groups of people who get good services.
- 164 there 's probably constants somewhere in here.
- 165 there 's times i go like to check her office see if she 's there.
- 166 there 's tip prevention sheets that come from the American Academy of Pediatrics.
- 167 there 's a variety of ways that you can think about motivating people.
- 168 there 's constant struggle and negotiation and some groups are dominant sometimes.
- 169 there are there are the other facts.
- 170 there 's times now i may just make things less clear all of a sudden.
- 171 there 're more all the time.
- 172 there are really four assumptions that are in that definition of natural selection.
- 173 there 're sometimes several species.
- 174 there 're a couple of people last week who forgot to give me their detailed outline.
- 175 there 's a lotta issues that we might not be considering.
- 176 there 's some Hebrew letters that um denote this spirit um the four spirits whatever.
- 177 there 's a there 's a there are more uh articles for your four thousand.
- 178 there are uh Francises on the one hand uh very much looking to the past.
- 179 there 's no sedative properties at all.
- 180 there 's no neighbors there 's no um chemical do you know what i mean?
- 181 there 's like five that are saved or something.
- 182 there 's actually lots of strategies that s- all seem plausible.

- 183 there 's other ways (we) gaining economy that that we obviously don't.
- 184 there 's a bunch of problems with that that all come to mind i me- in a kind of a jumble.
- 185 there are servants off to the side there 's a servant over here.
- 186 there 's a couple of ways of doing that.
- 187 there are three stars that make up the sword here.
- 188 there are forms that they don- p- there are animals supposedly that don't perceive but are aware.
- 189 there 's different people making it up this time so they might have a different style.
- 190 there are two and you can measure this receptor supersensitivity in many many many tissues.
- 191 there are only four four units of the airfield available.
- 192 there are two things that go into this discussion one is how many comparisons.
- 193 there 're several alternatives that 've been offered.
- 194 there are less extreme forms and so i i don't mean to throw out the baby with the bath water.
- 195 there 're seven or eight referees working on it another couple of people working on the computer aspect.
- 196 there 's all kind of things that can affect how you feel about a certain group of words.
- 197 there 's far too many.
- 198 there 're not just a finite number of arrangements but there 're an infinite number there.
- 199 there 's a few different ways you can change 'em um basically you 're normalizing them.
- 200 there are twenty thousand candidates right for burning today.
- 201 there 're other multi-level programs M-L-two etcetera is another one these.

- 202 there 's a lot of people that haven't danced before.
- 203 there are three kinds of reciprocity according to Marshall Sahlins.
- 204 there 's some downsides to retaking a course just from a purely academic U-of-M perspective.
- 205 there 's no there 's okay sorry there 's no objections.
- 206 there 're no collisions or anything of that sort we still have rho one one of V.
- 207 there 're Native Americans who speak Spanish but not English?
- 208 there are two parts to the exam.
- 209 there 's all sorts of problems with that.
- 210 there are slight defi- i mean there are slight differences between them.
- 211 there 're about a hundred chairs in this middle section.
- 212 there 's two booklets that we got.
- 213 there are uh tremendous conflicts going on about the establishment of Islam uh in uh societies.
- 214 there 's other side stairs that are n't those stairs that you came up.
- 215 there are three shuttles not two.
- 216 there 's two ways to get over to all these nodes.
- 217 there 's no more questions let 's thank Beth again and take a break.
- 218 there 's like five different communities i think in Detroit that they did interviews with.
- 219 there are no jobs
- 220 there are five languages on the reverse side of the card or you can get sign language if you want.
- 221 there 's a lot more um higher ratings in the in the yes category.
- 222 there 's more children using crack.

- 223 there 're in fact other T-cells uh within this region.
- 224 there 's semantic relations between words.
- 225 there 's so many little picky things that can go wrong and after you 've gotten everything put together.
- 226 there 're two general process models of Islamic revolutions in Africa or in the world.
- 227 there 's people from all over the world that come to America and they 're all citizens of the state.
- 228 there are uniform processes that occurred a long time ago and are still occurring and will continue.
- 229 there 's a lot of ups and downs in there.
- 230 there are tremendous increases in prison population um which has led to some decrease in crime.
- 231 there 's a whole bunch of 'em up in these those little apple tree.
- 232 there 's no there 's no hands and i mean it 's it 's linear it 's a lot like um i mean a lot of moder.
- 233 there 's scriptures in the New Testament that also like in Thessalonians and Phillipians.
- 234 there are some ways in which girls will always be girls.
- 235 there is there are things curved in my head.
- 236 there are kinda no more prisons to be built and no more people to be locked up.
- 237 there are some corrections that i would make but when it 's on i thought um it was it was on.
- 238 there 's there 's things that have two third- no quarks are the ones that have two thirds charge.
- 239 there 're all there 're all those B-flats that are the roots of four chords too.
- 240 there are eight people who only have friends basically who speak English.

- 241 there are other aspects and characters of leaves that you can use to determine climate.
- 242 there 's just like operons right?
- 243 there 's some people who you know don't perform very well on long written tests.
- 244 there 's some there are a couple things i was fuzzy on.
- 245 there 's more tributaries coming into it on the trellis.
- 246 there are more there so basically the cost overwhelmed the value to the consumers.
- 247 there are of us.
- 248 there 's four buttons and i 'll show you that in a minute.
- 249 there 's two pivotal concepts that i 'm going to examine today that is autonomy and the hospital environment.
- 250 there 're lots of things that are the same and and those things.
- 251 there are lots of examples out there just like this.
- 252 there 're plenty of people who think that but it doesn't really very often happen.
- 253 there 's quite a few and one of the first things was these serotonin selective drugs.
- 254 there 's okay sorry there 's no objections so i guess we can just we don't really have to vote.
- 255 there are so few i would add more.
- 256 there are decisions that why where postponing one year doesn't make me a whole lot better off.
- 257 there 's no states are n't there?
- 258 there 're various points of which i think i might wanna want clarification and or comment on.
- 259 there 's a bunch of extras here.
- 260 there 's a whole bunch a different ones.
- 261 there are some buttons lit up and each of these uh things has like a name on it.

- 262 there 's some areas where there 's tremendous loss of habitat.
- 263 there are additional factors that the immune system yes is one factor but there are other factors.
- 264 there 's nonmental objects environmental objects bodily objects nonben- nonmental objects.
- 265 there 's some cultures you don't say goodbye on the telephone.
- 266 there are zeros in the top row.
- 267 there 's.
- 268 there are lots of exceptions to this rule.
- 269 there are two down there i know but that wa- that was that was less yellow than a meadowlark.
- 270 there 're a lot of notes i hope you didn't look at all of them.
- 271 there 's also some some classrooms um for projection.
- 272 there 're just a few photons in the beam.
- 273 there are there are people i i i don't know like how they back it up.
- 274 there are lots of places on this campus where very critical academic work gets done.
- 275 there 's there 's some big issues there.
- 276 there are three types of Variant which are the same three types of the other Creutzfeld-Jakob disease.
- 277 there 's kind of a wide range of other languages and Arabic one other spoke Arabic.
- 278 there 're not many people there 's just overpopulation in in New York part of it.
- 279 there 's also graphical or written needs that people have.
- 280 there 's just a couple things i wanted to talk to you.
- 281 there 's a couple of questions i just wanna unravel then before we move to the next point.

- 282 there are a- there are social norms but then there 're also other factors that help people to choose.
- 283 there 's there 's six things that make up okay forget it maybe leptons are irreducible.
- 284 there 's other constructs which work perfectly well.
- 285 there 's also other uh things that fall under this heading.
- 286 there 's there 's example exams on the web or on the web page for two eighty-one.
- 287 there are two cats.
- 288 there 's sixteen of 'em.
- 289 there 's a lotta diseases that they can't cure they only can prevent with inoculation.
- 290 there 're an awful lot of churches where there are very far right and very far left people.
- 291 there 's not very many people that look at it from that viewpoint.
- 292 there are a couple things the the corner is i mean you could extrapolate it.
- 293 there 's varying degrees of autonomy and this is the one scale that captured that essence.
- 294 there 're lots of possibilities too.
- 295 there 's like four major ones right?
- 296 there are a couple of things that i wanna point out about these uh uh about these assumptions.
- 297 there are a number of interesting notational choices in this paper.
- 298 there 're all these babies now are emerging and they 're all (in their baby stage).
- 299 there are like flaws or like elements of like you know this is n't entirely like i don't know.
- 300 there 's more than one of you who shows up these are the hours they 're quite liberal.

- 301 there 's no other tools out there which measure organizational traits which support professional nursing.
- 302 there 're a few states where the whole state like Nebraska 's one where the whole state is full-day.
- 303 there are three different forms of the precipitate of mercury iodide.
- 304 there are five genes we 're only gonna talk about a couple of 'em.
- 305 there 're lots of different definitions of it um let 's see rather than graphic or than obvious.
- 306 there 's a lotta different flavors.
- 307 there are ones to study physical factors.
- 308 there 's two red cards and one black card.
- 309 there 's um say kids there 's tip prevention seep- sheets that come from the American Academy of Pediatrics.
- 310 there 's only three three possibly four bands but he 's in his fourth year.
- 311 there 's tandem repeats.
- 312 there 's two um actaeas and we 're only gonna take this one Actaea pachypoda.
- 313 there are several things that it 's really helpful to always try to get a grip on.
- 314 there 's at least fourteen different serotonin receptors that have been identified and cloned.
- 315 there 're different um phases of like virus tacking.
- 316 there are ways to correct for barometric pressure changes.
- 317 there are people that 've been in there six years and so there 're no beds for these guys.
- 318 there are people you guys haven't heard of probably much.
- 319 there are temporary exhibitions from our collections sometimes set up with a specific class.

- 320 there are enough important deists who do hold to a full-blown moral rationalism.
- 321 there 's lots a terms.
- 322 there 're so many (xx).
- 323 there 's a couple ways to go.
- 324 there 's i mean there 's not a lot of interventions out there C-B-C cent- um Center for Injury Prevention.
- 325 there 's a lot of bulbs and corms and underground stems rhizomes that have the same thing.
- 326 there 's a lot of components to that.
- 327 there are many different equations that have been put forth in the literature.
- 328 there are so many different kinds of so many different immigrant groups coming in.
- 329 there is some theoretical underpinnings.
- 330 there are obvious ones and indeed ones we are deeply committed to supporting like the library system.
- 331 there are theories (oh) i don't know if we 'll get there this semester.
- 332 there 're two things that we could think about that repetition does.
- 333 there 's three Cliff Swallows.
- 334 there 're two in the diagram i wish i had two of these things i should next time oh well.
- 335 there are is there a tra- there are travelling exhibits right now.
- 336 there are varieties of Spanish for example that have a bilabial fricative there.
- 337 there 's two basic dimensions.
- 338 there are difficulties in resolving it but there are there are voids all the way around the studio.
- 339 there are images in your mind or images in your brain or however you wanna put it.
- 340 there 's like a dozen of 'em.

- 341 there 's only five people in the class.
- 342 there 's multiple cats.
- 343 there are no orgasms there are no um sounds uh mental sounds uh uh um i- j- but there are states.
- 344 there are illusions.
- 345 there 's alcoves.
- 346 there are hundreds so you could picture a crossover at any locus mkay?
- 347 there 's two things to have a complete supply chain.
- 348 there 's also the peers.
- 349 there 's probably whole branches of philosophy that do nothing more than distinguish soul from mind.
- 350 there are no germinal centers and this would be viewed as normal.
- 351 there 's there 's two issues one is you don't have enough sampling points.
- 352 there are several types of deposits that are located throughout the basin.
- 353 there are people who are who are remarkable for what they stand for.
- 354 there are images with color and shape although there are no such things.
- 355 there 're bluebird houses around you find bluebirds.
- 356 there 's a couple different ones so there musta been a like a reporter in there.
- 357 there are lots of toothed leaves and they 're coming in.
- 358 there are two people only one can sign I-twenties but uh that 's Mary Coppler and Terri Mancini.
- 359 there is like all these other logical dependencies to these other principles coming from this base.
- 360 there 're multiple outbreaks of spatiality that are like disconnected from one another or something.

- 361 there are issues about the the assumptions that you 're making about your test.
- 362 there 's things that i 'm just feeling i 'm missing i 'll schedule a (new) right away.
- 363 there are a few spaces reserved for uh legacies you know people who are the children of big donors.
- 364 there are many popular putdowns of large women such as this postcard from the World War Two era.
- 365 there 're two models that are used frequently by people who are interested.
- 366 there are forty-two lectures just as many or more as in a regular term.
- 367 there are, there are many many more dead cells or dying cells within these germinal centers.
- 368 there 's a lot of different types of yeast.
- 369 there 's a lotta different educational issues.
- 370 there 're so many fun things to do.
- 371 there 's lo- there 's lots of best effort TAPs how does it decide which ones what priority to to do?
- 372 there 's extensive deposits of marine sedimentation which indicates that there were marine incursion.
- 373 there are two variations on drift the first is usually called the bottleneck effect.
- 374 there 's four hundred points in the class and lab counts one quarter of that.
- 375 there 's warm site conditions very rapid decomposition and mixing s- medium to high nutrients.
- 376 there are a lotta aspects of costs to an institution which are never (costed out) that is in dollars.
- 377 there 's two checks.
- 378 there are divorces or deaths or uh intense interpersonal experiences in the family.
- 379 there are probabilities.

- 380 there 's leisure and income and we let working correspond to leisure-equal-to-one.
- 381 there 're a lot of men oh they 're fine we don't need to study them.
- 382 there are six units of it and we were only working in a small part of it.
- 383 there 's efforts to preserve it.
- 384 there 's like many here that 's why i could make oh okay.
- 385 there 're a variety of definitions about of watersheds or of a drainage basin.
- 386 there are facts of subjective experience that i have a the the there 's a fact about my experience.
- 387 there 're like ten of those.
- 388 there are two other things that i wanted to say.
- 389 there 're i think five themes that he concentrated on that would be like significant in the future.
- 390 there are structural motivations that you can try to (f-) analyze.
- 391 there are these sociological problems etcetera.
- 392 there are a lot that probably stay very close to home.
- 393 there 're still serotonin receptors in the brain.
- 394 there 's three things that are important to keep in mind.
- 395 there are cases.
- 396 there are a number of changes that are unrelated to each other.
- 397 there 's more than one so if you just in effect forget about the jus-.
- 398 there are things curved in my head.
- 399 there are many journals and i get most of them.
- 400 there are multiple voices.
- 401 there are um women who have to you know they have have children or you know single mothers.

- 402 there are large and easy-to-use buttons on this device.
- 403 there 's a lot of questions that this kind of film brings up.
- 404 there are people that think that about America.
- 405 there are characteristics is a necessity for there being a contradiction.
- 406 there are are large scale projects going on um funded by a lot of sources.
- 407 there are a lot of bands that can keep up putting out schlock for years and years and years.
- 408 there are.
- 409 there 's a lot of terms for the same thing.
- 410 there 's guns guns.
- 411 there 's other things that are affecting those things.
- 412 there is also there 're specific articles in the Universal Declaration of Human Rights which defends these.
- 413 there 's some others like uh egret plumes.
- 414 there 's no complex numbers in this so that 's kinda weird.
- 415 there 's two of 'em.
- 416 there 's all all different types of validity and we talked about a few of 'em uh f-about a month.
- 417 there 's different levels of organization you can look at.
- 418 there are many days throughout the week that women can't play golf that only men are allowed.
- 419 there 's really two groups of questions or at least we have considered two groups of questions.
- 420 there 's times when you need less of some product.
- 421 there 're a lot of subgroups in the Senufo as well.
- 422 there are forty-seven ronin the forty-seven samurai.

- 423 there 's bound to be more than one of those that you 'd really wanna do at a time.
- 424 there 's also issues that we 're not gonna get into too much.
- 425 there 's some internal validity issues.
- 426 there 's a lot of things in there that are n't quite making it.
- 427 there 're four experiments.
- 428 there 's a hell of a lot of different ways to do it.
- 429 there 's some environmental programs or something we can cut to give more treatment.
- 430 there 's three trues you multiply them.
- 431 there are just rules for how you have to behave.
- 432 there 's two of each.
- 433 there are sort of several obvious facts so take take the second one unified consciousness.
- 434 there are some other particular references there 's one to two Kings chapter one about a-.
- 435 there are proteins and eukaryotic R-N-A polymerase that look just like them.
- 436 there 's a lotta pigmentation changes that 're going on.
- 437 there 's DNA viruses R-N-A viruses and retroviruses.
- 438 there 're a number of studies of that type.
- 439 there are some people who think that access really access to resources shouldn't even be equitable.
- 440 there are a few who get awards for for a semester.
- 441 there 's only two possible explanations God or chance so there must be a God.
- 442 there are also people working on systems where you put in the telegraphic speech.
- 443 there 're some points of unclarity just jump in an- and you know yell and let me know.

- 444 there are a lot of sexual expectations.
- 445 there are strains of animals that have an inherently defective immune system.
- 446 there 's twelve kinds of human understanding which are called twelve categories of understanding.
- 447 there are two piles it 's the exact same handout pick one up on your way out.
- 448 there are correlations that help us estimate diffusion.
- 449 there are other body fluids there 's the lymphatic system there 's the cerebral spinal fluid.
- 450 there 're infinitely many ways.
- 451 there 's two strategies when they 're developing products.
- 452 there 're actually two promoters there is P-one and P-two.
- 453 there 's a hundred to four hundred and fifty cases of it.
- 454 there are two minus-tens two minus-thirty-fives.
- 455 there are so few women scientists and engineers my main goal has been to kind of fit in.
- 456 there 're no distortions externalities taxes uh we should be getting this maximized subject.
- 457 there are fifteen degree zones and that represents one hour of movement.
- 458 there 're two types of cone- connectors.
- 459 there are also less obvious ones that i think we also need to think about as public goods.
- 460 there 's multiples of something but if you give one all you need is one so it doesn't matter.
- 461 there are transitional forms.
- 462 there are people that take mica sheets and um-.
- 463 there are several different possible explanations for this.

- 464 there are two okay then we t-.
- 465 there are two distinct in comparing these sets of lines with this line.
- 466 there 's twelve or thirteen eggs laid.
- 467 there are two equations.
- 468 there 's lots of different themes thematic things.
- 469 there are strippers in the world.
- 470 there are people that especially there 's a certain kind of a a Buddhist.
- 471 there 're two reasons that i selected the Great Basin okay.
- 472 there 's a lot of factors that you have to consider but uh i i do think we can be judicious.
- 473 there 're special CSP sections for each.
- 474 there 's some offices in there which are kind of temporary three-quarter partition sorts things.
- 475 there are two homeworks due uh no homework one is due in two weeks.
- 476 there are three math questions and then the fourth question.
- 477 there are equivalent sentence elements.
- 478 there 're other places where i did the same thing i mean so now i just wanna clarify.
- 479 there 's just tons of historical examples where you sort of know something.
- 480 there 's different levels of organization you can look at in humans.
- 481 there are all of these offices in listed under that.
- 482 there are two like crabapple trees in that pine.
- 483 there are legal and regulatory battle uh barriers to advanced practice.
- 484 there are some countries that you know have signed and they ignore the agreement.
- 485 there 's different species that have gotten it b- you know.

- 486 there are people who say we don't have an image illusion.
- 487 there 's like really close up shots of like genitalia.
- 488 there 's states that that s- sell special license plates.
- 489 there 're actually other descendants as well.
- 490 there are loopholes in the different ways that they 're being cheated.
- 491 there 's hunting accidents.
- 492 there 's like different degrees of hot.
- 493 there 's people who come in and sit in the back instead of going to the seats.
- 494 there are proteins being phosphorylated here by this binding event.
- 495 there 's different ways of you you combine strategies to to get by.
- 496 there 's eight masses and you know and uh and so you can kinda get along.
- 497 there 're at least four new audiences that Peking opera went after.
- 498 there 's a lot of strong women in the Bible.
- 499 there 's two different things for creative writing versus English if you wanted to be a professor.
- 500 there 's almost none now.