A Corpus-Based Study of Variation and Change in Adverb Placement Across World Englishes

A Thesis submitted to the Faculty of the Graduate School of Arts and Sciences of Georgetown University in partial fulfillment of the requirements for the degree of Master of Science in Linguistics

By

Edwin Ko, B.A.

Washington, DC May 5, 2016
A CORPUS-BASED STUDY OF VARIATION AND CHANGE IN ADVERB PLACEMENT ACROSS WORLD ENGLISHERES

Edwin Ko, B.A.

Thesis Advisor: Amir Zeldes, Ph.D

ABSTRACT

Recent investigations have shown an effect of geography on syntactic variation among dialects in a handful of languages. As dialect corpora are becoming increasingly available, developments in corpus-based dialectology have introduced sophisticated analysis techniques for studying dialect differences. For example, in a study analyzing variation in adverb position, spatial autocorrelation techniques have been employed to identify regional clusters from a corpus of written Standard American English (Grieve 2012b). In another study, Euclidean distances were used to measure morphosyntactic differences in British dialects from the Freiburg Corpus of English Dialects (FRED, Szmrecsanyi 2011). However, not much is known about syntactic variation among socially and geographically non-contiguous varieties such as World Englishes, and the effectiveness of modern dialectometric techniques in these types of studies.

In this study, adverb position is analyzed for geographical patterning across ten varieties of spoken and written English from the International Corpus of English (ICE) – Northern Ireland, Republic of Ireland, Kenya, Tanzania, Jamaica, Hong Kong, India, Philippines, Singapore, and Canada. Three adverb positions: sentence-initial, preverbal, and postverbal, and four adverb classes: evaluative, frequency/temporal, manner, and modal adverbs, were chosen and their probabilities were extracted from each corpus. This study then employs a number of statistical analysis and visualization techniques to identify similarities and differences due to geographical space.
The analysis shows significant overall spatial clustering for five variables in the spoken data: frequency/temporal adverbs in postverbal position, manner adverbs in preverbal and postverbal positions, and modal adverbs also in preverbal and postverbal positions. None of the variables in the written texts were found to exhibit significant geographical patterning. In the spoken data, ‘hot spots’ were analyzed around Ireland and Canada for postverbal position, and around India for placing adverbs before the verb. In all of these societies, bilingualism and multilingualism are integrally interwoven, and many speakers learn English as their second-language. Possible explanations for variation in adverb placement are intense contact with the languages of these locales and similarities to placement preferences in English when it was first introduced in these locations as a result of colonialism.

INDEX WORDS: Adverb placement, World Englishes, syntactic variation, corpus linguistics, geolinguistics, language contact, multilingualism
This thesis would not have been possible without the support of many people. First, I would like to thank my advisor, Amir Zeldes, and my two committee members, Mark Sicoli and David Lightfoot, for the countless hours of fruitful discussions, and for their support throughout my graduate studies. I am also indebted to Cathy O’Connor who has instilled within me a passion for research. Finally, none of this would not have been possible without the unwavering support from my family and my friends, and for that, I thank them.
# Table of Contents

## Chapter

1. Introduction .................................................. 1
2. Background & Related Work ................................. 4
   2.1 Adverb Placement ........................................ 4
   2.2 Syntactic Variation ..................................... 11
   2.3 Corpus-based Dialectology ............................ 14
3. Spatial Statistics ............................................. 19
   3.1 Mantel Test .............................................. 19
   3.2 Global Moran’s I ......................................... 21
   3.3 Local Getis-Ord Gi* ................................... 23
4. Data & Methodology ........................................... 25
   4.1 International Corpus of English ........................ 25
   4.2 Geographic Data ......................................... 26
   4.3 Corpus Preprocessing ................................... 27
   4.4 Adverb Class ............................................ 27
   4.5 Corpus Analysis ......................................... 31
   4.6 Spatial Analysis ......................................... 34
5. Results .......................................................... 39
   5.1 Mantel Test .............................................. 39
   5.2 Global Moran’s I and Local Getis-Ord Gi* ............ 40
6. Discussion ....................................................... 49
   6.1 Effects of Contact Languages ........................... 52
   6.2 Similarities to Historic Englishes ..................... 55
   6.3 Probabilistic Knowledge and Indigenization .......... 57
   6.4 Diachronic Syntax ....................................... 58
   6.5 Limitations ............................................... 60
7. Conclusion ....................................................... 65

## Appendix

A. Appendix: List of Adverbs ................................. 67
B. Appendix: Adverb Placement Frequencies from the Spoken Texts ... 68
C  Appendix: Adverb Placement Frequencies from the Written Texts . . . 70
D  Appendix: Results for a Distance Band of 3000km . . . . . . . . . . . . 72
E  Appendix: Results for a Distance Band of 6000km . . . . . . . . . . . . 79
F  Appendix: Results for a Distance Band of 8550km . . . . . . . . . . . . 86
List of Figures

4.1 Distribution of adverbs in sentence-initial position (left), preverbal position (center), and postverbal position (right) within the spoken texts ................................................................. 33
4.2 Distribution of adverbs in sentence-initial position (left), preverbal position (center), and postverbal position (right) within the written texts ................................................................. 34
5.1 Relative frequencies in frequency/temporal adverbs in preverbal position. 42
5.2 Local Getis-Ord Gi* values for frequency/temporal adverbs in preverbal position. ................................................................. 43
5.3 Relative frequencies for frequency/temporal adverbs in postverbal position. ................................................................. 43
5.4 Local Getis-Ord Gi* values for frequency/temporal adverbs in postverbal position. ................................................................. 44
5.5 Relative frequencies for manner adverbs in preverbal position. ........ 44
5.6 Local Getis-Ord Gi* values for manner adverbs in preverbal position. 45
5.7 Relative frequencies for manner adverbs in postverbal position. ...... 45
5.8 Local Getis-Ord Gi* values for manner adverbs in postverbal position. 46
5.9 Relative frequencies for modal adverbs in sentence-initial position. ... 46
5.10 Local Getis-Ord Gi* values for modal adverbs in sentence-initial position. 47
5.11 Relative frequencies for modal adverbs in preverbal position. ....... 47
5.12 Local Getis-Ord Gi* values for modal adverbs in preverbal position. 48
D.1 Local Getis-Ord Gi* values for evaluative adverbs in sentence-initial position. ............................................. 72
D.2 Local Getis-Ord Gi* values for evaluative adverbs in preverbal position. 73
D.3 Local Getis-Ord Gi* values for evaluative adverbs in postverbal position. 73
D.4 Local Getis-Ord Gi* values for frequency/temporal adverbs in sentence-initial position. ............................................. 74
D.5 Local Getis-Ord Gi* values for frequency/temporal adverbs in preverbal position. ............................................. 74
D.6 Local Getis-Ord Gi* values for frequency/temporal adverbs in postverbal position. ............................................. 75
D.7 Local Getis-Ord Gi* values for manner adverbs in sentence-initial position. ............................................. 75
D.8 Local Getis-Ord Gi* values for manner adverbs in preverbal position. 76
D.9 Local Getis-Ord Gi* values for manner adverbs in postverbal position. 76
D.10 Local Getis-Ord Gi* values for modal adverbs in sentence-initial position. 77
D.11 Local Getis-Ord Gi* values for modal adverbs in preverbal position. 77
D.12 Local Getis-Ord Gi* values for modal adverbs in postverbal position. 78
E.1 Local Getis-Ord Gi* values for evaluative adverbs in sentence-initial position. ............................................. 79
E.2 Local Getis-Ord Gi* values for evaluative adverbs in preverbal position. 80
E.3 Local Getis-Ord Gi* values for evaluative adverbs in postverbal position. 80
E.4 Local Getis-Ord Gi* values for frequency/temporal adverbs in sentence-initial position. ............................................. 81
E.5 Local Getis-Ord Gi* values for frequency/temporal adverbs in preverbal position. ............................................. 81
E.6 Local Getis-Ord Gi* values for frequency/temporal adverbs in postverbal position. .......................................................... 82
E.7 Local Getis-Ord Gi* values for manner adverbs in sentence-initial position. .......................................................... 82
E.8 Local Getis-Ord Gi* values for manner adverbs in preverbal position. 83
E.9 Local Getis-Ord Gi* values for manner adverbs in postverbal position. 83
E.10 Local Getis-Ord Gi* values for modal adverbs in sentence-initial position. 84
E.11 Local Getis-Ord Gi* values for modal adverbs in preverbal position. 84
E.12 Local Getis-Ord Gi* values for modal adverbs in postverbal position. 85
F.1 Local Getis-Ord Gi* values for evaluative adverbs in sentence-initial position. .......................................................... 86
F.2 Local Getis-Ord Gi* values for evaluative adverbs in preverbal position. 87
F.3 Local Getis-Ord Gi* values for evaluative adverbs in postverbal position. 87
F.4 Local Getis-Ord Gi* values for frequency/temporal adverbs in sentence-initial position. .......................................................... 88
F.5 Local Getis-Ord Gi* values for frequency/temporal adverbs in preverbal position. .......................................................... 88
F.6 Local Getis-Ord Gi* values for frequency/temporal adverbs in postverbal position. .......................................................... 89
F.7 Local Getis-Ord Gi* values for manner adverbs in sentence-initial position. .......................................................... 89
F.8 Local Getis-Ord Gi* values for manner adverbs in preverbal position. 90
F.9 Local Getis-Ord Gi* values for manner adverbs in postverbal position. 90
F.10 Local Getis-Ord Gi* values for modal adverbs in sentence-initial position. 91
F.11 Local Getis-Ord Gi* values for modal adverbs in preverbal position. 91
F.12 Local Getis-Ord Gi* values for modal adverbs in postverbal position. 92
List of Tables

4.1 Geographic information of the locations during corpus compilation . . 26
6.1 Year of initial colonization for each location (in ascending longitude) . 56
B.1 Placement counts and percentages for evaluative adverbs in the spoken texts ................................................................. 68
B.2 Placement counts and percentages for frequency/temporal adverbs in the spoken texts .................................................... 68
B.3 Placement counts and percentages for manner adverbs in the spoken texts ................................................................. 69
B.4 Placement counts and percentages for modal adverbs in the spoken texts 69
C.1 Placement counts and percentages for evaluative adverbs in the written texts ................................................................. 70
C.2 Placement counts and percentages for frequency/temporal adverbs in the written texts .................................................... 70
C.3 Placement counts and percentages for manner adverbs in the written texts ................................................................. 71
C.4 Placement counts and percentages for modal adverbs in the written texts 71
Chapter 1

Introduction

Tobler’s (1970) first law of geography is a widely known observation in geography that states, “Everything is related to everything else, but near things are more related than distant things.” This phenomenon has also been observed by geneticists, going by the name of Isolation by Distance (Sewall 1943). In dialectology, Nerbonne and Kleiwig (2007) term this the Fundamental Dialectological Postulate: geographically proximate dialect varieties tend to be more similar than distant ones.

Many researchers have observed an influence of geographical distance on phonological, lexical and syntactic variation in dialects from a handful of languages by exploring regional patterns (e.g. North American English: Kurath 1949, 1972, Labov et al. 2006, Grieve 2012b; British English: Goebl 2007, Szmrecsanyi 2012; and Dutch: Spruit 2006, 2008, Spruit et al. 2009). As dialect corpora are becoming increasingly available, developments in corpus-based dialectology have introduced sophisticated analysis techniques for studying dialect differences.

In particular, spatial autocorrelation techniques are used to identify spatial clustering of certain linguistic features. These methods have been employed in studies in geography for several decades. However, they have only recently been introduced to studies in dialectometry, a sub-field of dialectology. This field stemmed from the seminal work by Sèguy in 1971, that embraces these kinds of computational and statistical tools to study linguistic variation due to geographical space.
Still not much is known about linguistic variation among varieties of English spoken around the world, which are geographically and socially non-contiguous. Travel between countries such as Singapore, India and the Philippines, where English is mainly spoken as a second language, is a non-trivial task, and most people do not have the resources to do so. This places restrictions on the interaction of speakers between these locations. Therefore, there does not appear to be any reason to suspect that geographical space has a role to play in similarities or differences among World Englishes. So if differences among World Englishes are found, what are the likely mechanisms underlying the variation?

This thesis investigates the effect of geography on adverb placement across varieties of English worldwide using corpus-based dialectometric methodology. In English, several adverbs may occupy several positions in a given utterance without any discernible change in meaning. For example, consider the sentences in (1) and (2) containing the adverbs *loudly* and *often*, respectively, in preverbal and postverbal positions.

(1)  a. Logan *loudly* sings.
    b. Logan sings *loudly*.

(2)  a. Logan *often* sings.
    b. Logan sings *often*.

By employing a corpus-based approach, it is possible to analyze naturalistic spoken and written data collected from hundreds of speakers rather than depending on interviews of just a few speakers. This provides a better representation of the whole speech community. Data come from the International Corpus of English (see http://ice-corpora.net/ice) which includes ten varieties of English: Northern Ireland, Republic of Ireland, Canada, Jamaica, Hong Kong, Singapore, India, the Philippines, Kenya
and Tanzania, all of which have been colonized in the past resulting in the prominent use of English today in these locations. Furthermore, bilingualism and multilingualism are integrally interwoven within these societies and so are in intense contact with other languages. If systematic geographical patterns arise, then we may be able to predict similar properties in other varieties of English.

This thesis is organized in the following way. Section 2 discusses the relevant background and previous works including classification schemes of adverbs that have previously been proposed, syntactic variation and the issues of semantic equivalence, and corpus-based dialectology. In Section 3, I describe three spatial statistical techniques that are used in this study: the Mantel test, global Moran’s I, and local Getis-Ord Gi*. Section 4 discusses the data used in this study, and outlines the methodology. Section 5 reports the results which find significant geographical patterning of adverb placement within the spoken texts, but not in the written texts, while Section 6 discusses the importance of these findings and possible explanations as to why these patterns arise. Section 7 concludes.
Chapter 2

Background & Related Work

2.1 Adverb Placement

The syntax of adverbs has traditionally been a notably complex domain of research. In relation to other constituents and syntactic elements, adverbs tend to be optional and they may occupy different syntactic positions. The surface positions that adverbs may appear in can even vary. There may also be restrictions in where adverbs can be placed, and the semantic properties of these adverbs are oftentimes sensitive to their placement within a sentence.

The distribution of adverbs and its complexity have been a constant refrain within the literature. In his seminal work, Jackendoff (1972: 47) once identified adverbs as “the least studied and most misaligned part of speech...maltreated beyond the call of duty.” In 1984, Vendler (as cited in Waters 2011) recalls a piece of advice from Zellig S. Harris: “Don’t try adverbs, that way madness lies,” and in the first chapter of his 2002 book *The Syntax of Adjuncts*, Thomas Ernst remarks, “nobody seems to know what to do with adverbs.” To this day, the study of adverbs remains controversial, and Ernst (2014: 1) notes that, “If the study of adverbs once represented a “swamp”, as was sometimes remarked forty years ago, it is perhaps now low-lying marshy ground, messy and prone to flooding, but with a few solid paths meandering though it.” It is clear then that one should tread lightly when dealing with what appears to be the chaos of adverbs.
Two approaches are generally associated with the classification of adverbs. The first approach analyzes adverbs based on their positional distribution (Jackendoff 1972, Alexiadou 1997, Cinque 1999), and the second approach classifies adverbs on account of their semantic properties (Ernst 2002).

Jackendoff’s (1972) study on adverb placement identifies three main adverb positions: sentence-initial, preverbal (or auxiliary) position, and postverbal (or clause-final) position, and formulates two adverb classes based on their permissible placements: S adverbs and VP adverbs. S adverbs, which consist of speaker- and subject-oriented adverbs, can appear sentence-initially, directly after the subject, or immediately following the first auxiliary:

(3) (Unfortunately,) Logan (unfortunately) had (unfortunately) been hit by a car.

In contrast, VP adverbs, which include manner adverbs, can appear after the verb or directly before the verb:

(4) Taylor (loudly) sang (loudly).

Potsdam (1998) suggests an additional adverb class, E(xtent) adverbs, which may only appear clause-internally:

(5) (Potsdam 1998, Ex. 27c, 29c)
   a. They (hardly) should (hardly) worry about that.
   b. *Hardly they should worry about that.

More recent syntactic analyses of adverb placement have been referred to as “specifier-based” approaches (e.g. Alexiadou 1997, Cinque 1999). According to Cinque (1999), there exists a universal hierarchy of functional projections in which adverbs are generated in the specifier position of these projections.
The order of the functional projections (commonly referred to as Cinque’s Hierarchy) is displayed in (8), which is hypothesized to hold cross-linguistically. The hierarchy is based on a fixed relative order of adverbs.\(^1\) To illustrate this, the following sequence of adverbs in Italian and French are argued to be in a fixed relative order:

(6) (Cinque 1999: 7, Ex. 22)

a. solitamente > mica > già > più > sempre > completamente (Italian)
b. généralement > pas > déjà > plus > toujours > complètement (French)

According to Cinque, sequences of adverbs that violate this order will result in an ill-formed sentence:

(7) (Cinque 1999: 7, Ex. 21)

a. Jean a toujours complètement perdu la tête pour elle.
b. *Jean a complètement toujours perdu la tête pour elle.

Each adverb in a given order thus defines a class of adverbs, and each adverb class forms its own functional projection. Therefore, adverbs belonging to a specific adverb class would then appear in the specifier of the corresponding projection. For example in (8), frankly belongs to the adverb class, \(\text{Mood}_{\text{speech act}}\). Cinque (1999: 84) suggests that other adverbs that fall within this class include honestly and sincerely. Similarly, fortunately falls within the class, \(\text{Mood}_{\text{evaluative}}\), and other adverbs belonging to this class include unfortunately, luckily, regrettably, surprisingly, strangely/oddly (enough), and (un)expectedly (Cinque 1999: 85).

(8) Cinque’s Hierarchy (1999: 106, adapted)

\[
\begin{array}{c}
\text{[frankly } \text{Mood}_{\text{speech act}}]\ 
\text{[fortunately } \text{Mood}_{\text{evaluative}}]\ 
\text{[allegedly } \text{Mood}_{\text{evidential}}]\ 
\text{[probably } \text{Mod}_{\text{epistemic}}]\ 
\text{[once } T(\text{Past})]\ 
\text{[then } T(\text{future})]\ 
\text{[perhaps } \text{Mood}_{\text{irrealis}}]
\end{array}
\]

\(^1\)Adverbs appearing outside these fixed positions is said to be the result of movement.
The heads of these projections may contain verbal morphology (e.g. auxiliaries and verbal suffixes), which share a tight semantic relation with the adverb. Because of the strong semantic ties between adverbs and verbal heads, adverbs are said to enter in a spec-head agreement within their projections. Furthermore, the verbal morphology is claimed to also be in a fixed relative order, similar to that of the ordering of adverbs, and Cinque (1999: 106) argues that adverbial specifiers and their (verbal) heads “match in a systematically one-to-one fashion.” The relative ordering of verbal morphology (as the functional heads) in tandem with the ordering of adverbs (as the specifiers) thus provide additional support for this hierarchy.

\[ \text{necessarily } \text{Mod}_{\text{necessity}} \ [\text{possibly } \text{Mod}_{\text{possibility}} \ [\text{usually } \text{Asp}_{\text{habitual}} \ [\text{again } \text{Asp}_{\text{repetitive(I)}} \ [\text{often } \text{Asp}_{\text{frequentative(I)}} \ [\text{intentionally } \text{Mod}_{\text{volitional}} \ [\text{quickly } \text{Asp}_{\text{celerative(I)}} \ [\text{already } \text{T}(\text{Anterior}) \ [\text{no longer } \text{Asp}_{\text{terminative}} \ [\text{still } \text{Asp}_{\text{continuative}} \ [\text{always } \text{Asp}_{\text{perfect}} \ [\text{just } \text{Asp}_{\text{retrospective}} \ [\text{soon } \text{Asp}_{\text{proximative}} \ [\text{briefly } \text{Asp}_{\text{durative}} \ [\text{characteristically } \text{Asp}_{\text{generic/progressive}} \ [\text{almost } \text{Asp}_{\text{prospective}} \ [\text{completely } \text{Asp}_{\text{SyCompletive(I)}} \ [\text{tutto} (\text{Italian}) \ [\text{Asp}_{\text{PlCompletive}} \ [\text{well Voice} \ [\text{fast/early } \text{Asp}_{\text{celerative(II)}} \ [\text{again } \text{Asp}_{\text{repetitive(II)}} \ [\text{often } \text{Asp}_{\text{frequentative(II)}} \ [\text{completely } \text{Asp}_{\text{SyCompletive(II)}} \]

---

2 According to Cinque, "tutto" (Italian), which means ‘all’ or ‘everything’, appears in the specifier of the projection AspPlCompletive. However, no such element exists in English (see Cinque (1999) for further discussion).

3 Several adverb classes may appear in a different order. For example, the (Italian) “frequentative” adverb spesso ‘often’ may appear after or before the “repetitive” adverb già ‘again’:

\( (8) \) (Cinque 1999: 92, Ex. 42)

a. (Quando troviamo qualcosa) questa è spesso già stata scoperta da qualcuno.
   ‘(When we find something) this has often already been discovered by someone.’

b. Questa proprietà è già stata scoperta spesso, negli ultimi cinquant’anni.
   ‘This property has already been discovered often, in the last fifty years.’

The two distinct positions correspond to specific scope positions, and the same lexical items may appear in either position. A projection such as Asp_{frequentative(I)} occupies a position that is higher than Asp_{frequentative(II)}.
In contrast, Ernst (2002) takes a scope-based account towards a theory of adverbs, and employs a classification, displayed in (9), based on the adverb’s lexicosemantic properties. According to the classification, adverbs are divided into three main groups: predicational, domain, and functional. These groups are then arranged into subclasses. For example, among predicational adverbs, speaker-oriented adverbs express the attitude of the speaker whereas subject-oriented adverbs reflect the properties of the subject. Again, these adverbs are further divided into even more fine-grained categories. Subject-oriented adverbs consist of agent-oriented and mental-attitude adverbs, the latter of which modify the event as well as the agent and the former describes “a state of mind experienced by the referent of the subject of the verb” (Ernst 2002: 63). These lexicosemantic properties also act as requirements that dictate which positions adverbs may appear in based on the so-called Fact-Event-Object Calculus, which is a “set of rules for binding events and propositions” (Ernst 2002: 50). However, the precise details regarding his proposal are beyond the scope of this thesis.

(9) Ernst (2002: 9, 96, 327):
A. Predicational:
   a. speaker-oriented:
      i. speech-act: frankly, briefly, simply
      ii. evaluative: oddly, amazingly, predictably
      iii. epistemic: modal: probably; evidential: obviously
   b. subject-oriented:
      i. agent-oriented: rudely, tactfully, wisely
      ii. mental-attitude: calmly, willingly, intentionally
   c. exocomparative: similarly, accordingly, independently
   d. aspect-manner: slowly, quickly, abruptly
   e. pure-manner: tightly, loudly, woodenly
B. Domain: mathematically, chemically
C. Functional:
   a. negation: not, never
   b. focusing/clausal-degree: even, just, only
c. time-related:
   i. location-time: now, once, at noon
   ii. duration: long, briefly, the whole day
   iii. aspectual: still, soon, yet
   iv. frequency: sometimes, always, frequently

d. quantificational:
   i. frequency\(^4\)
   ii. habitual: generally, usually, habitually
   iii. additive: again

e. clausal relations: purpose, clausal, concessive, conditional, etc.

One of the challenges in capturing the distribution of adverbs is that the placement of an adverb can influence its meaning. Consider the following sentences:

(10) (Ernst 2002: 42, Ex. 2.4, emphasis added)

a. Alice *cleverly* has answered the questions.

b. Alice has *cleverly* answered the questions.

c. Alice has answered the questions *cleverly*.

In (10a), there is a single meaning of *cleverly*, in that Alice was clever for answering the question (i.e. a subject-oriented reading). In (10c), there is also a single meaning of *cleverly*, but the meaning is different from (10a). In the sentence (10c), the adverb has a manner reading and can be paraphrased as *Alice has answered the questions in a clever way*. According to Ernst, the sentence in (10b) is ambiguous between the readings found in (10a) and (10c).

There are also adverbs that can occupy different positions, but with subtler nuances in their meaning. Examples include aspect-manner adverbs such as *quickly*:

\(^4\)As Ernst (2002: 327) notes, “frequency adverbs have both temporal and quantificational characteristics,” and therefore appear among time-related adverbs as well as quantificational adverbs.
(11) Ernst (2002: 85, Ex. 2.149, emphasis added)

a. Lynn quickly raced down the hallway.

b. Lynn raced down the hallway quickly.

In (11a), *quickly* is used to describe the speed at which the event had begun, but in (11b), it describes the speed (or manner) of the action. Other examples include manner adverbs that can have a reading of a so-called resultative adverb (Geuder 2000) which includes adverbs such as *heavily*, *elegantly*, and *beautifully*:

(12) Geuder (2000: 34, Ex. 40, emphasis added)

a. Mary *elegantly* dressed (only manner)

b. Mary dressed *elegantly* (resultative or manner)

The sentence (12a) can only have a manner interpretation whereby Mary got dressed in an elegant way. In (12b), there are two possible readings. The first is that of a manner reading while the alternative describes the result of the action, that is roughly, Mary got dressed and as a result of getting dressed, Mary looked elegant.

Adverbs that do not appear to change in meaning have also been discussed within the literature. Jackendoff (1972: 42) notes that adverbs such as *quickly*, *slowly*, *reluctantly*, *sadly*, *quietly*, *indolently*, *frequently*, *immediately*, *often*, and *soon* do not have a “discernible change in meaning.”

Furthermore, Ernst notes that pure manner adverbs are unable to obtain a clausal reading, and thus can only have a manner reading:

---

5 For some of these adverbs, I argue that there is a change in meaning:

(13) a. Sadly, Taylor lost the tournament.

b. Taylor lost the tournament *sadly*.

In (13a), *sadly* is a speaker-oriented (or evaluative) adverb and can be paraphrased as *it’s sad that Taylor lost the tournament*. In contrast, the adverb in (13b) receives a manner reading.
The shift in adverb placement therefore does not appear to alter its meaning. A trait that is shared among pure manner adverbs is that they “often require an event specified as involving a physical stimulus, such as volume of sound” (Ernst 2002: 87). Pure manner adverbs thus include *loudly, brightly, limply, tightly, woodenly,* and other adverbs that fit this criterion while preserving its meaning when occupying a different position.

The study of adverbs has played a significant role in many areas of generative syntax: adverbs are widely used as probes (Potsdam 1998); the distribution of adverbs is usually employed as a diagnosis for verb-movement (Pollock 1989) and other flavors of movement such as C-to-I lowering in Irish (McCloskey 1996); adverbs provide important insights on object-shift and scrambling (Thráinsson 2008), on functional projections (Cinque 1999), among other syntactic phenomena. Their optionality offers a more holistic and in-depth understanding of syntactic structures, and their variable placements within a clause make adverbs natural candidates for investigations on syntactic variation.

2.2 Syntactic Variation

One hallmark of human language is the vast amount of variability that it exhibits. In studying syntactic variation, generativists tend to be interested in ‘competence’, or our knowledge of grammaticality, and employ introspection and grammaticality judgments as their main source of data. On the other hand, variationists look towards
‘performance’, or usage, in order to describe a person’s linguistic behavior by discovering quantitative regularities in their language use. As Lavendera (1978) puts it, the former is a theory of sentences and the latter is a theory of utterances. The view that ‘grammar is grammar and usage is usage’ (Newmeyer 2003) has been commonplace since the beginning of transformational generative grammar:

“Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-communication, who know its (the speech community’s) language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of this language in actual performance.”

(Chomsky 1965: 3)

Within the generative tradition, Principles and Parameters (P&P) is an approach (Chomsky 1981) that seeks to distinguish the fundamental principles that are invariant in human language, and to identify the set of macro- and micro-parameters which determine in what ways languages can vary. Followers of this approach believe that these universal principles and parameters are part of the innate properties of Universal Grammar (UG), and thus P&P is a theory of not only UG, but also a theory of cross-linguistic, and more recently, the concepts adopted from P&P have been prevalent in research on cross-dialectal syntactic variation (see Zanuttini and Horn 2014, and Adger and Trousdale 2007 for recent works on dialect syntax).  

The view among variationists regarding syntactic variation remains controversial. In variationist studies, the sociolinguistic variable describes how two or more surface

---

6While the P&P framework is no longer considered part of mainstream generative syntax, many concepts are still highly influential within the Minimalist Program (MP) (Chomsky 1995). This is not surprising given that MP developed out of P&P.
constructions of a single underlying variable are conditioned by internal, linguistic factors as well as external ones such as age, sex, and class (Tagliomonte 2006). These variants are considered to be “alternative ways of saying the same thing” (Labov 2008: 42). The question that has been the source of this ongoing debate is whether syntactic variants can ever be semantically equivalent. But what exactly does it mean for two or more variants to be semantically equivalent? As noted by Tagliamonte (2006: 73), “in the case of variables functioning at the level of discourse or pragmatics, the notion of semantic equivalence becomes even more problematic.” Thus in the strongest view, semantic equivalence is unattainable since utterances encode extralinguistic information such as a speaker’s attitudes towards the hearer that would differ among individuals and among contexts (Bolinger 1977).

There has been considerable resistance to extending the sociolinguistic variable to include syntactic variation due to operational and conceptual reasons (Lavendera 1978; Romaine 1981, 1984; Cheshire 1987). Nevertheless, there have been significant studies within this enterprise (actives/passives, Weiner and Labov 1983; copula deletion, Rickford 1998; zero complementizer, Tagliamonte and Smith 2005, among others). According to Weiner and Labov (1983: 30), they assume that active and passive constructions “normally have the same meaning in a truth-conditional sense” and by restricting the term meaning, they acknowledge that they are only using a “rough semantic equivalence.” Under this assumption, they found little influence from social factors in the choice of the agentless passive suggesting that syntactic variants tend not to have much social significance in comparison to phonological variants. Moreover, Sankoff (1988: 153) introduced the notion of neutralization in discourse which states that “distinctions in referential value or grammatical function among difference surface forms can be neutralized in discourse.” The requirement of semantic equivalence therefore depends on the way ‘meaning’ is construed in these kinds of studies.
The assumption of meaning equivalence between syntactic variants has also contributed to important findings in psycholinguistics on genitive alternation (Rosenbach 2002), particle placement (Gries 2003), and dative shift (Bresnan 2007). Romaine (1984: 416) has suggested that “the relevant external factors in [the choice of syntactic variant] is what one might call a pragmatic constraint,” and these are precisely the constraints that are shown to play a significant role in the processing strategies involved. Therefore, one would expect processing and contextual cues such as information structure, animacy and grammatical weight to influence the choice of syntactic variant. These contextual weights would then allow us to predict a speaker’s syntactic choice given a certain context.

Differences in the underlying probabilistic constraints were found in cross-varietal studies of the three syntactic alternations mentioned above – i.e. dative, genitive, and particle placement. For example, Australians tend to prefer prepositional datives over Americans in dative alternation (Bresnan and Ford 2010), and the length of the possessum is significant for American, but not for British English in genitive alternation (Hinrichs and Szmrecsanyi 2007). More recently, Szmrecsanyi et al. (2016) found that the very same predictors were significant in several English varieties across the world; the only differences were in their effect sizes. This therefore suggests that the choice in syntactic variant in these constructions is governed primarily by pragmatic and discourse considerations as opposed to social constraints.

2.3 Corpus-based Dialectology

Traditionally, dialect maps and atlases draw from surveys and questionnaires. But these data have been considered to be too noisy, limited, and imprecise for identifying regional variation which required “looking for overall trends” (Grieve 2016: 98).
For the most part, these kinds of analyses were highly subjective as it depended on the judgments of the dialectologist, and therefore could not be replicated. The work by Séguy (1971) had been seminal in introducing the methodology known as dialectometry which, according to Szmrecsanyi (2011: 45), is “the branch of geolinguistics concerned with measuring, visualising and analysing aggregate dialect similarities or distances as a function of properties of geographic space” through the use of computational and quantitative techniques. Thus, the notion of ‘corpus-based’ is simply the integration of and dependence on naturalistic corpus data, and the marriage between corpus linguistic methodology and statistical techniques. Like in sociolinguistics and psycholinguistics, there is a “concern with nonstandard dialects, use of non-introspective data and statistical methodology” (Sankoff 1989: 146).

While individual linguistic features (or variables) have been investigated in dialectology to distinguish dialects, it has been argued that studies incorporating geolinguistic signals such as geographical distance should instead aggregate multiple features for a more accurate characterization (Nerbonne 2009; Szmrecsanyi 2008, 2011, 2014). First, single features may not truly be representative of the dialects in question and second, investigations looking at other features may contradict previous findings. Recent studies have used dialectometric approaches with multiple features to distinguish areal patterns and dialect regions which corroborates existing hypotheses (Dutch dialects: Spruit 2006, 2008; American English dialects: Grieve et al. 2011, 2013; British English dialects: Szmrecsanyi 2008). Other studies have even looked at non-contiguous varieties like World Englishes, and by aggregating morphosyntactic features of several varieties, they attempted to characterize dialects holistically in

---

7It is important to mention that Smrecsanyi is not advocating multiple-feature studies across-the-board. He notes that “[individual features] are fine when the research question is tree-centered – i.e. when it is the features themselves that are of analytic interest” (Smrecsanyi 2014: 91).
order to view typological differences from a bird’s-eye view (Szmrecsanyi and Kortmann 2009, Kortmann and Szmrecsanyi 2011). The push towards aggregating linguistic features has been attributed to the combination of novel computational and quantitative techniques and the increased availability of dialect corpora.

There are many statistical techniques that have been employed to identify linguistic differences (or distances) among dialects. In studies that aggregate features, a common method is to calculate the overall distance between pairs of dialects, instead of comparing features one-by-one as was done in the past. The Euclidean distance measure, which is based on the Pythagorean theorem, calculates the distance of two dialects by taking into account the distances of every feature that the dialects share. The formula used to calculate these distances is shown in Equation 1:

\[
d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_n - y_n)^2} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}
\]

where \(k\) is the number of features, \(x_i\) is the relative frequency of feature \(i\) of dialect \(x\), and \(y_i\) is the relative frequency of feature \(i\) of dialect \(y\) (Szmrecsanyi 2011: 566-567).\(^8\) The Euclidean distances between each pair of dialects as a (linguistic) distance matrix can then be analyzed in a number of ways: directly plotting these distance values using cartographic visualizations (e.g. beam maps), correlating the linguistic distance matrix with other distance matrices (e.g. as-the-crow-flies distances), or through other means. For example, drawing from the Freiburg English Dialect Corpus (FRED), Szmrecsanyi (2012) analyzes three distances matrices: as-the-crow-flies distance, travel cost (in terms of travel time), and Trudgill’s (1974) notion of linguistic gravity (TGLI) which incorporates geographical distances with population statistics, and

\(^8\)Other distance measures include Hamming distance (see Spruit 2008 for syntactic distance) or Levenshtein distance (see Nerbonne and Kleiweg 2007 for phonetic distance).
to explain the Euclidean morphosyntactic distances from 57 morphosyntactic features of several British English dialects. Using regression analysis, he finds that morphosyntactic distance, while significantly correlated with all three geography-related distance matrices, is mostly influenced by TGLI, which explains 24% of the variance.

Other methods include spatial analysis techniques that test whether an individual feature exhibits any geographical patterning, as opposed to aggregating features which offers no insight into how influential an individual feature may be. In other words, the distance measures do not explain “if an individual linguistic variable exhibits a regional pattern, in the way that logistic regression allows for social patterns to be identified in the distribution of a linguistic variable” (Grieve 2012a). In his work, Grieve (2012b) employs two spatial autocorrelation techniques: Global Moran’s I and Local Getis-Ord Gi*, in order to identify whether placements of certain adverbs exhibit any regional patterning overall and whether there are specific locations that act as ‘hot spots’ for patterning in written Standard American English. In one study, he analyzes ‘common’ adverbs (see the appendix of Grieve 2012b for the list of ‘common’ adverbs) and adverbs longer than five characters that end with -ly for three adverb position variables: infinitive splitting (e.g. to loudly sing), non-modal auxiliary splitting (e.g. has loudly sung), and modal auxiliary splitting (e.g. can loudly sing). In another study, he analyzes two positions: sentence-initial and sentence-internal/final, for four adverb class variables: temporal adverbs, however, also, and instead. In these studies, significant geographical patterning was found for one position: modal splitting, and for two adverb classes: however and also. These results are interesting because many of the variables consistently identified the dialects of the Northeast against those in the Southeast and South Central states. Furthermore, these findings were based on written registers which had been claimed to conceal much of the author’s speech patterns (Schneider 2002).
Spatial patterns offer a unique window into the effects of a variety of linguistic and biological phenomena including language contact, language spread through space and time, language and human diversity, and population movement (Nichols 1992). Language variation is intimately interwoven with language change and so geostatistical techniques that are able to reveal underlying spatial patterns may also reflect the outcomes of language change and allow us to understand the deeper histories of these diachronic processes (Haynie 2014). Corpus-based dialectometry therefore generates many productive lines of inquiry which may also span historical linguistics.
Spatial statistics have been widely applied in fields such as geography (Miller 2004), ecology (Fortin et al. 2002), and population genetics (Guillot et al. 2009). One of most widely used techniques is spatial autocorrelation which measures clustering or dispersion due to geographical space. A significant positive spatial autocorrelation represents clustering of a feature, and a significant negative spatial autocorrelation indicates dispersion. Within the past decade, the dialectometric enterprise has embraced the use of spatial autocorrelation techniques for identifying underlying spatial patterns of individual linguistic features as well as aggregates of features (e.g. Nerbonne and Kleiwig 2007, Grieve et al. 2011). In this section, I outline three spatial autocorrelation methods that are employed in this study: the Mantel test, global Moran’s I, and local Getis-Ord Gi*.

3.1 Mantel Test

The Mantel (1967) test evaluates the correlation between two symmetrical distance matrices using the Pearson’s product-moment correlation coefficient where the values of the matrices are not necessarily independent to each other. Distance matrices commonly found in dialectometric studies are typically a matrix of a specific linguistic feature correlated against a geographical distance matrix. These distance matrices can be calculated in a variety of ways including the Euclidean distance measure discussed
in Section 2.3. The Mantel test statistic is based on the cross-product of the two distance matrices displayed in Equation 2 (Mantel 1967).

\[
(2)
\]

\[
z = \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ij}y_{ij}
\]

where \( n \) is the number of elements in each of the symmetrical matrices, \( x \) and \( y \) each correspond to one of the two distance matrices, and \( x_{ij} \) and \( y_{ij} \) are the distance values measured for location \( i \) and \( j \).

This statistic measures the significance of the correlation between the two matrices through repeated permutations (McCune and Grace 2002). First, the test measures the correlation between the two original matrices. Then, the matrices are permuted and the test is re-applied to the matrices containing the new arrangements. The process of permutation and re-application of the statistic is repeated \( P \) number of times, where \( P \) is given as an input parameter for the desired number of repetitions. In general, as the number of permutations increases, so too does the precision of the p-value. The p-value is obtained by comparing the original result (from the non-randomized matrices) to the distribution of results from the permutations. By normalizing the Mantel statistic, a regular Pearson correlation coefficient (\( r \)) can be obtained. The formula for normalization is shown in Equation 3.

\[
(3)
\]

\[
r = \frac{1}{n - 1} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{(x_{ij} - \bar{x})(y_{ij} - \bar{y})}{s_x s_y}
\]

To illustrate how one might interpret these results, let matrix \( x \) be the matrix of linguistic distances, and matrix \( y \) be a matrix of as-the-crow-flies distances. A significant positive correlation would therefore indicate that as the linguistic distance increases, the geographical distance between pairs of locations does too. On the other hand, a
significant negative correlation indicates that as the linguistic distance decreases, the
geographical distance also decreases (Grieve 2014). ¹

3.2 Global Moran’s I

Global Moran’s I (Moran 1950) is a measure to detect whether a variable exhibits
significant positive or negative global spatial autocorrelation. Specifically, it identifies
any significant spatial clustering or dispersion for any given variable from a bird’s eye
perspective. It is used to characterize correlations among feature values due to space.
The formula for global Moran’s I is displayed in Equation 4.

\[
I = \frac{N \sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}
\]

where \( N \) is the total number of locations, \( x_i \) is the value of the feature in location (or
dialect) \( i \), \( \bar{x} \) is the mean of the values of the feature across all locations, and \( w_{ij} \) is
the value of the spatial weighting function, or the spatial weight between location \( i \)
and location \( j \) determined by some function (Grieve et al. 2011, Grieve 2012a, 2012b,
2014). Essentially, the statistic compares the difference between the value of a feature
at the target location and the mean value of all features to the difference between the
value of the feature in each neighboring location and the mean value of all features.

The spatial weighting function allows pairs of locations that are closer together
to receive more weight than locations that are farther apart. Two main weighting
functions can be used: a reciprocal weighting function and a binary weighting func-
tion.² The former assigns weights based on the reciprocal of the distance between a
pair of locations. In this way, the weighting decreases as the distance between two
locations increases. The latter assigns a binary weighting based on a cutoff distance.

¹For potential limitations of the Mantel test in dialectometric studies, see Grieve 2014.
²The combination of the reciprocal and binary weighting function is also sometimes used.
If the pair of locations is equal to or within the cutoff distance, a weighting of 1 is assigned, whereas locations outside the cutoff distance will receive a weighting of 0. For example, if the assigned cutoff distance was 100 miles, then $w_{ij} = 1$ if $\text{distance}_{ij}$ (i.e. the distance between location $i$ and $j$) $\leq$ 100 miles, and $w_{ij} = 0$ if $\text{distance}_{ij} > 100$ miles. Thus, features outside the cutoff do not significantly influence features located within the cutoff.\(^3\)

The calculated Moran’s I value falls within $[-1, 1]$, in which a significant negative value represents an overall dispersion – i.e. the neighbors of a feature with a high value will tend to have low values. Similarly, the neighbors of a feature with a low value will tend to have high values. A significant positive value represents an overall clustering whereby features with high values tend to have neighboring features that also have high values, and vice versa. In other words, a significant positive value indicates that features tend to have similar values to nearby features. A standardized z-score is used to determine the significance of the Moran’s I value if it is equal to or larger than the corresponding level of significance. If more than one variable is being analyzed and compared, then the significance level is adjusted using a Bonferroni correction. In a study comparing five variables, the z-score must be equal to or larger than 0.233 corresponding to a one-tailed significance level of 0.01, adjusted using a Bonferroni correction ($0.05/5 = 0.01$). According to Grieve et al. (2011), a one-tail (as opposed to a two-tailed) significance level is used if the aim of the study is to determine whether

\(^3\)As Grieve et al. (2011: 12) remarks, “The cutoff distance essentially sets the level of resolution for the analysis. A smaller cutoff is better for identifying smaller clusters, whereas a larger cutoff is better for identifying larger clusters. Setting the cutoff distance is problematic, however, because it is possible for different linguistic variables to exhibit regional patterns at different levels.” Furthermore, Grieve et al. (2011) suggest calculating a variety of different cutoffs, and then selecting the cutoff that exhibits the most significant spatial autocorrelation.
there is significant positive global spatial autocorrelation (or clustering) in the data, as opposed to also detecting negative global spatial autocorrelation (or dispersion).

3.3 Local Getis-Ord Gi*

Local Getis-Ord Gi* (Ord and Getis 1995) is a measure to detect which locations exhibit significant high or low value spatial autocorrelation. In other words, it identifies spatially significant locations or regions by looking at where features with high or low values cluster. If a location were to be identified as a high value hot spot for a particular variable, then nearby locations tend to also display high values for that variable. Global Moran’s I differs from local Getis-Ord Gi* in that the former detects a significant spatial patterning within the whole area of study, whereas the latter focuses on individual features and their spatial distribution by detecting significant high and low value hot spots. Unlike global Moran’s I which returns a single value for a given variable, local Getis-Ord Gi* returns a value for each location. The formula involved in calculating the local Getis-Ord Gi* statistic is defined in Equation 5.

\[
G_i^* = \frac{\sum_j w_{ij} x_j - \bar{x} \sum_j w_{ij}}{\sqrt{\sum_j x_j^2}} - \bar{x} \sqrt{N \sum_i w_{ij} - (\sum_j w_{ij})^2}
\]

where \( N \) is the total number of locations, \( x_j \) is the value of the feature in location (or dialect) \( j \), \( \bar{x} \) is the mean of feature values across all locations, and \( w_{ij} \) is the value of the spatial weighting function, or the spatial weight between location \( i \) and location \( j \) determined by some function (Ord and Getis 1995, Grieve et al. 2011, Grieve 2012a, 2012b, 2014). In short, the statistic compares local averages – i.e. the mean of the feature’s value at the target location and the feature’s values of neighboring locations, to global averages. In order to be significant, neighboring locations of a hot spot will have values that tend to correspond to the value of the hot spot.
The statistic returns a z-score for each location which is then used to determine which locations act as significant hot spots for high or low value clustering. Statistical significance is determined by comparing the z-score to the corresponding significance level. Like global Moran’s I, if more than one variable is being analyzed and compared, then the significance level is adjusted using a Bonferroni correction. However, it is important to note that unlike global Moran’s I which uses a one-tailed test in dialectometric studies, local Getis-Ord Gi* employs a two-tailed test because the aim is to identify high and low value hot spots (Grieve 2014). A statistically significant positive z-score represents a clustering of high values at the particular location, while a significant negative z-score represents a clustering of low values, where higher or lower z-scores signify greater intensity.
Chapter 4

Data & Methodology

This section first describes the corpus data used in the study, which comes from the International Corpus of English (ICE) (see http://ice-corpora.net/ice), and the geographic data (i.e. geographic coordinates and population statistics), which is obtained from a variety of government and institutional websites. The rest of the section is then dedicated to the methodology employed in this study, which includes corpus cleaning and prepossessing, the choice of linguistic variables (i.e. adverb classes and syntactic position), a report of the placement frequencies, and the application of spatial analysis techniques described in the previous sections.

4.1 International Corpus of English

The ICE corpus consists of sub-corpora of over ten different varieties of English spoken worldwide. In this study, I analyze corpora of ten varieties of World English: ICE-Ireland (Northern Ireland and the Republic of Ireland), ICE-East Africa (Kenya and Tanzania), ICE-Jamaica, ICE-Hong Kong, ICE-India, ICE-Philippines, ICE-Singapore, and ICE-Canada. Each corpus contains approximately a million words (600,000 words in the spoken texts and 400,000 words in the written), and they follow a similar corpus design in order to restrict the amount of variation due to genre differences (Nelson 1996).\textsuperscript{1} It is important to note that while the Irish and East African

\textsuperscript{1}In their attempt to follow a common corpus design, several corpus compilers expressed difficulty in obtaining certain kinds of spoken and written texts such as parliamentary
corpora contain a million words, these corpora have been divided so that each variety – i.e. Northern Irish English (or Ulster English), Irish English (or Hiberno-English), Kenyan English and Tanzanian English, end up containing 500,000 words in total for both spoken and written registers.

4.2 Geographic Data

Geographic information such as longitudes and latitudes was obtained from NASA’s Latitude/Longitude Finder (2015) by entering the name of the location. The population statistics during the years the corpus were constructed had been obtained from The World Bank (2015):

<table>
<thead>
<tr>
<th>Location</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Population</th>
<th>Date of Compilation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada</td>
<td>56.130366</td>
<td>-106.346771</td>
<td>27,790,000</td>
<td>1989-2007</td>
</tr>
<tr>
<td>Jamaica</td>
<td>18.109581</td>
<td>-77.297508</td>
<td>2,390,000</td>
<td>1994-2008</td>
</tr>
<tr>
<td>Northern Ireland</td>
<td>54.787715</td>
<td>-6.492315</td>
<td>1,596,000</td>
<td>1990-1994</td>
</tr>
<tr>
<td>Republic of Ireland</td>
<td>53.41291</td>
<td>-8.24389</td>
<td>3,514,000</td>
<td>1990-1994</td>
</tr>
<tr>
<td>Tanzania</td>
<td>-6.369028</td>
<td>34.888822</td>
<td>25,480,000</td>
<td>1990-1996</td>
</tr>
<tr>
<td>Kenya</td>
<td>-0.023559</td>
<td>37.906193</td>
<td>23,446,000</td>
<td>1990-1996</td>
</tr>
<tr>
<td>India</td>
<td>20.593684</td>
<td>78.96288</td>
<td>868,900,000</td>
<td>post-1989</td>
</tr>
<tr>
<td>Singapore</td>
<td>1.352083</td>
<td>103.819836</td>
<td>3,047,000</td>
<td>post-1989</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>22.396428</td>
<td>114.109497</td>
<td>5,704,000</td>
<td>post-1989</td>
</tr>
<tr>
<td>Philippines</td>
<td>12.879721</td>
<td>121.774017</td>
<td>61,950,000</td>
<td>post-1989</td>
</tr>
</tbody>
</table>

Table 4.1: Geographic information of the locations from during corpus compilation (in ascending longitude).

The population statistics were obtained during the years the corpus was being built in order to better reflect the actual populations at the time.\(^2\)

\(^2\)The years during which the corpora were constructed were obtained from the manuals accompanying each of the the corpora, if available. For those that did not include this
4.3 Corpus Preprocessing

Several preprocessing steps were involved to facilitate the extraction of placement counts. First, the spoken and written components of each corpus were separated in order to study the differences in variation between the two registers. The spoken texts were then cleaned by removing tags containing metalinguistic information such as pauses, unclear words, or uncertain transcriptions. Then they were sentence and string tokenized, and tagged for their parts-of-speech using the Python library package, NLTK Toolkit (Bird et al. 2009).

4.4 Adverb Class

In this study, I investigate three adverb positions: sentence-initial, preverbal (i.e. between the subject and verb), and postverbal positions, and four adverb classes: evaluative, frequency/temporal, manner, and modal adverbs.

It is generally agreed among scholars that adverbs can change in meaning when placed in a different syntactic position. As Ernst (2002) remarks, a class of adverbs that do not change in meaning is the so-called pure manner adverbs such as loudly.

(15) a. Logan sang loudly.
    b. Logan loudly sang.
    c. Loudly, Logan sang.

Semantically, there is no difference in truth-conditional meaning between the three sentences in (15).\(^3\) Therefore, the position in which speakers place these adverbs is not driven by the information they intend to convey as is the case for adverbs that

\(^3\)While there is no difference in meaning, there may very well be discourse-pragmatic or functional reasons as to why an adverb may be placed in one position over another.
do change in meaning. However, pure manner adverbs are not commonly used in spoken and written language and suffer the effects of data sparsity, especially within the spoken texts. After counting all tokens tagged as adverbs across all corpora, six of the most common pure manner adverbs, loudly, quietly, tightly, loosely, gently, and brightly, were only found seven times in the spoken corpora of Canadian and Tanzanian English, whereas within the written Singaporean English texts, there were 51 occurrences. These frequencies considered adverbs across all syntactic positions and so the counts would have been much lower after distributing them into the three positions analyzed in this study. Therefore, I expand the class of manner adverbs to include aspect-manner adverbs (Ernst 2002) – e.g., quickly, slowly, rapidly, and resultative adverbs (Geuder 2000) – e.g., heavily, beautifully, elegantly.

4It is important to mention that these semantic classes contain some overlap. For example, Ernst (2002) lists tightly as an example of a pure-manner adverb, but I argue that it functions similarly to resultative adverbs as described by Geuder (2000).

(16) a. Jordan tightly sealed the lid.
    b. Jordan sealed the lid tightly.

In (16a), there is only a manner reading in which Jordan sealed the lid in a tight way, but in (16b), it is ambiguous between a manner and resultative interpretation whereby the latter describes that Jordan sealed lid so that it resulted in a tight lid.

(17) a. Taylor went to the store quickly.
    b. Taylor quickly went to the store.

(18) a. Hayden sang beautifully.
    b. Hayden beautifully sang.

According to Ernst (2002), the sentence in (17a) can only receive a manner reading describing the action and its celerity, but in (17b), the sentence is ambiguous between a manner reading and a reading that describes the alacrity of the event. Furthermore, Geuder (2000) indicates that beautifully acts as a resultative adverb in that (18a) is
ambiguous between a manner and a resultative interpretation – i.e. Hayden sang in a way that reflected their physical beauty upon singing.

The limitation of extending manner adverbs to include aspect-manner and resultative adverbs is clear. If the meaning of the adverb can differ based on its placement within a sentence, then the position of these adverb is dependent on the speaker’s intended meaning. This should then be a serious issue. However, there are several important considerations here: (a) the differences in meaning among aspect-manner and resultative adverbs are very subtle and may be prone to being neutralization in discourse (Sankoff 1989), and (b) the locales analyzed in this study are all bi- and multilingual societies and for many, English is spoken as a second language so speakers may be subject to syntactic transfer effects. Because working with a corpus removes the investigator from the actual speakers, there is no way to tell what they had originally intended to say. Nevertheless, I assume that these adverbs, for the most part, have very little to no difference in meaning. This issue will be expanded further in Chapter 6.

According to Ernst (2002), modal adverbs reflect the degree of the speaker’s beliefs about the truth value of the proposition.

(19) Ernst (2002: 73, Ex. 2.108, emphasis added)

a. Sam has *probably* made an appointment.

b. *Sam has made an appointment *probably*.

Given the sentences in (19), modal adverbs can only occur preverbally but not postverbally. However, modal adverbs can freely appear postverbally as parenthetics (or with a comma intonation) (Hübler 1983). Because these adverbs reflect the speaker’s assertions about the truth of the proposition, I consider modal adverbs to not change in meaning.

29
Another adverb class that was chosen for this thesis is FREQUENCY/TEMPORAL. I argue that habitual adverbs (see p.6, 5.C.d.ii.) should also have temporal and quantification characteristics, just as frequency adverbs do (Ernst 2002: 327). According to Ernst (2002: 350), habitual adverbs are a subclass of frequency. The main difference between habitual and frequency adverbs is the greater tendency for habituals to occur higher in the clause than regular frequency adverbs. Thus, if habitual adverbs are a subset of frequency adverbs, then it should follow that habitual adverbs appear among time-related and quantificational adverbs.

(20) Ernst (2002: 349, Ex. 7.114, emphasis added)

a. Carol frequently was buying gifts.

b. Carol was frequently buying gifts.

c. Carol was buying gifts frequently.

(21) a. Tyler had went to Montana recently.

b. Recently, Tyler had went to Montana.

(22) a. Tyler went to Montana recently.

b. Recently, Tyler went to Montana.

According to Ernst, in (20a), there are multiple gift buying events, whereas in (20b), there is only one event of multiple gift buyings; (20c) is ambiguous between the two interpretations. Furthermore, in (21b), there are two readings. In one reading, recently acts as the reference point from which it is considered Tyler went to Montana, and in the other reading, it is the point in time at which Tyler went to Montana. In (21a), the reading is only that of the latter.\(^{5}\) It is not clear if these distinctions are carried over to (22) in which the tense and aspect differ. Regardless, these nuances are very subtle, and so perhaps in discourse, these nuances may also be neutralized.

\(^{5}\)More intuitive examples might involve substituting recently for last night.
Lastly, I consider the class of pure evaluative adverbs which express the speaker’s evaluation of the state of affairs (Ernst 2002). Pure evaluatives differ from other evaluatives in that the former only functions as an evaluative, whereas the latter can also differ in meaning.

(23) Ernst (2002: 76, Ex. 2.123, emphasis added)
   a. *Oddly, Carol was dancing.
   b. Carol was dancing oddly.

(24) Ernst (2002: 76, Ex. 2.121, emphasis added)
   a. *Unfortunately, she lay down on a scorpion’s nest.
   b. *She lay down unfortunately.

In (23a), the most natural reading is an evaluative one, although a manner interpretation is also possible if emphasis is placed on oddly. In (23b), a manner reading is only possible, but with a comma intonation, we can also get an evaluative reading. On the other hand, pure evaluative adverbs can only have an evaluative interpretation like in sentence (24a). Similar to modal adverbs, evaluatives may not appear postverbally unless as parentheticals. Therefore, I adopt pure evaluatives as an adverb class in this study, and like modals, I assume that evaluatives do not change in meaning.

4.5 Corpus Analysis

All tokens tagged as adverbs were extracted from across all the spoken and written corpora using a Python script. Following the discussion in the previous section, the most common adverbs were then manually selected based on the examples provided for each adverb class – evaluative, frequency/temporal, manner, and modal – by Cinque (1999), Geuder (2000) and Ernst (2002), and also based on my own
judgments (see Appendix for the list of adverbs chosen in this study). Then, by parsing through each corpus, if a token matched one of the adverbs within the four classes, there are several criteria that determined their placement within the sentence.

The placement counts (or frequencies) were extracted from the spoken and written texts separately to determine where differences between the two registers exist. Nevertheless, the process by which the frequencies were extracted from both types of texts was identical. Extracting placement counts for SENTENCE-INITIAL position was a relatively easy process. The corpora were sentence tokenized and each sentence occupied its own line in the text files. Because of this, if the first token of the string contained an uppercase character as its first character, then a count was added to the overall sum for sentence-initial position of the class that the adverb belongs to. However, discourse markers such as so, um and uh may precede adverbs that would otherwise be in sentence-initial position. Therefore, an adverb that directly follows a discourse marker or a sequence of discourse markers would still be considered in sentence-initial position. An adverb is in POSTVERBAL position if (a) it is the last token in the sentence, (b) the preceding token was tagged as a verb and the following token is not a verb, (c) the two preceding tokens consisted of a verb and a modifier (e.g. very), respectively, and (d) the next token is either a preposition, conjunction, or punctuation. If an adverb did not meet the criteria for either sentence-initial or postverbal position, then it is in PREVERBAL position. One hundred sentences in each corpus were checked manually for correct placement of adverbs, and of the 1,000 sentences across all corpora, the Python script achieved more than 95% accuracy.

For each adverb class, the relative frequency of each position was calculated following the formula in Equation 6.
Relative Frequency = \frac{C_i}{C_{initial} + C_{preverbal} + C_{postverbal}}

where \( C_i \) is the count for adverb position \( i \), and the denominator represents the total count for all positions (initial, preverbal, and postverbal) for a particular adverb class. (See Appendix for the raw counts and relative frequencies, expressed as a percentage, of each variable across all locations).

Figure 4.1: Distribution of adverbs in sentence-initial position (left), preverbal position (center), and postverbal position (right) within the spoken texts.

Figure 4.1 displays the distribution of adverbs and their placements across all locations for the spoken data. Sentence-initial position is generally disfavored across all adverb classes, especially manner adverbs which almost categorically does not appear sentence-initially. Manner adverbs in pre- and postverbal position also appear to have similar distributions. Furthermore, preverbal position is a common placement among adverbs with the exception of evaluative adverbs, which is placed in all three positions at around the same rate (in comparison to the other adverb classes), with a preference for postverbal position.
Figure 4.2: Distribution of adverbs in sentence-initial position (left), preverbal position (center), and postverbal position (right) within the written texts.

Figure 4.2 displays the distribution of adverbs and their placements across all locations for the written data. Surprisingly, the distribution of adverbs in the written texts appears to be very similar to the distribution of adverbs in the spoken texts. Therefore, there does not seem to be any noticeable difference in how adverbs are placed between the written and spoken registers.

4.6 Spatial Analysis

The relative frequencies for each variable were incorporated into three spatial statistics that were used to measure both global and local spatial autocorrelation.

4.6.1 Mantel Test

The Mantel test measures the association between two distance matrices – i.e. syntactic distance correlated with a geographical distance matrix. Syntactic distances were calculated by measuring the difference in relative frequencies between each pair of locations and placed in a matrix for each adverb class. Therefore, for each adverb class, the Euclidean distance between location A and location B would be the sum of the squared difference for sentence-initial, preverbal, and postverbal position (see Equation 1). The values were then placed in a matrix.
The two geographic distance measures chosen in this study were as-the-crow-flies distance, which is simply the straight-line distance between point A and point B (without taking into consideration terrains such as mountain ranges or oceans and other bodies of water), and Trugill’s (1974) notion of linguistic gravity, a measure that takes into account both geographical distance and population size. The as-the-crow-flies distance between two locations was calculated by measuring the distance between the geographic coordinates, obtained from the NASA Latitude/Longitude Finder (2015), of pairs of locations. These values were then entered into a matrix as one of the geographical distance matrices.

The diffusion of linguistic innovations in the gravity model proposed by Trudgill (1974) is portrayed as a wave whereby one population center affects nearby smaller population centers. These smaller populations would act as centers of gravity themselves, influencing even smaller population centers nearby (Nerbonne and Heeringa 2007). This model is related to the model of gravity in physics; while the moon and its rotation is mainly affected by the larger, neighboring Earth, the sun’s mass is mainly influencing the Earth’s rotation and thus is only indirectly associated with the moon’s rotation. Within this model, social contact, measured by geographical closeness, and contact frequencies, measured by population size, play a vital role. The equation for Trudgill’s Linguistic Gravity Index (TGLI) is calculated using the formula in Equation 7, adopted from Nerbonne and Heeringa (2007).

\[
I_{ij} = \log \frac{P_i P_j}{(d_{ij})^2}
\]

where \(I_{ij}\) represents the logged mutual influence between the two population centers, \(P_i\) and \(P_j\) are the populations at location \(i\) and \(j\), and \(d_{ij}\) is the distance between locations \(i\) and \(j\). The idea is that that interaction between two populations will correlate
positively as distance between them decreases, but correlate negatively as distance increases. Furthermore, the equation postulates that the greater the two populations, the more social contact.\textsuperscript{6} By incorporating population statistics, obtained from the World Bank (2015), and as-the-crow-flies distances, the linguistic gravity values for each pair of locations were then calculated and placed in another geographic distance matrix.

For each adverb class, the Euclidean syntactic distance of each pair of locations were correlated with the as-the-crow-flies distances, and also with the linguistic gravity distances. The spoken and written data were analyzed separately, and a total of 9,999 permutations was performed for each computation of the Mantel test.\textsuperscript{7}

4.6.2 Global Moran’s I and Local Getis-Ord Gi*

A binary spatial weighting function was used with a cutoff distance of 6,000km although a cutoff distance of 3,000km and 8,550km were tested as well (see Appendix for the results of each of these cutoff distances). A 6,000km cutoff was selected because it provided the optimal resolution for this study. A cutoff of 3,000km was considered too small because it resulted in three locations without any links to neighboring locations. Therefore, these locations were only compared with the overall averages of the values and were not compared to the averages of their neighbors as is the case for locations that do have links.

\textsuperscript{6}Asymmetric population centers, in which one population is considerably larger than the other, are not accounted for by the model. Rather, only the product of the two populations is taken into account. Therefore, two population centers with two million people each will share the same linguistic gravity as two population centers, where one has a million people and the other has four million people, given that the distances between the pairs of locations are identical.

\textsuperscript{7}The Mantel test statistic (mantel.rtest) in the R library \{ade4\} was used to measure the correlation between the two distance matrices.
While the significance of the results was more robust for a 6,000km cutoff, there are also theoretical motivations for this distance band over the larger 8,550km one. The average number of links for the 6,000km cutoff is 2.6, and for 8,550km, the average number of links is 4.6. The 8,550km cutoff includes pairs of locations such as Kenya and Singapore, and Ireland and Jamaica, two pairs which would not have been included if a 6,000km cutoff was used instead. The choice of a smaller cutoff therefore ensures that locations that are farther away and do not appear to have any social or cultural ties such as between Kenya and Singapore are not included in the comparisons. Having said that, locations that are not within the cutoff, but may be related due to factors such as migration or colonization (e.g. British rule of Hong Kong) will not be included. As Keith Ord suggested, via personal correspondence, non-distance links can also be determined by considering cultural factors that would “use binary links and assign connections only to pairs of countries that had a common cultural bond.” While this could be incorporated in determining links of ‘related’ locations, I focus only on geographical distance and leave this for future research.

In order to determine the significance of global Moran’s I, the p-value must be less than or equal to 0.016, adjusted using Bonferroni correction \((0.05/3 = 0.016)\) (or greater than or equal to a corresponding z-score of 2.14 for a one-tailed p-value as in this case because the interest here is whether there is any spatial clustering).\(^8\) Since the alternative hypothesis was specified via the parameter, conversion from the z-score value to its p-value correspondent is not required and so the p-value can be directly compared. For local Getis-Ord Gi*, the p-value is also 0.016. Since we are interested in whether a location exhibits local spatial autocorrelation of low or high values, we are interested in a two-tailed p-value, which corresponds to a z-score of

\(^8\)The global Moran’s I (moran.test) and local Getis-Ord Gi* (localG) test statistics used in the study were from the \{spdep\} R library.
±2.41. In contrast to the global Moran’s I which returns only one p-value, the local Getis-Ord Gi* returns a z-score for each location.
This section reports the results for the spatial autocorrelation statistics: Mantel test, global Moran’s I and local Getis-Ord Gi*, for twelve different variables (four adverb classes and three adverb positions) in both spoken and written registers. The Mantel test was used to measure the correlation between syntactic distances and two geographical distances, as-the-crow-flies and Trudgill’s (1974) notion of linguistic gravity (TGLI). Global Moran’s I tests whether there is significant overall clustering, and local Getis-Ord Gi* tests which locations act as hot spots for high or low value clustering. The z-scores from the local Getis-Ord Gi* test is plotted on a map. The relative frequencies across the ten locations are also plotted on a map and displayed alongside the results for global Moran’s I and local Getis-Ord Gi*. Due to the large large number of results for the three spatial statistics – 16 variables for the Mantel test (four adverb classes and two registers), and 72 for global Moran’s I and local Getis-Ord Gi* (three cutoff distances, four adverb classes, three adverb positions, and two registers), only the significant results are displayed here for a cutoff of 6,000km (see Appendix for the results for the global Moran’s I and local Getis-Ord Gi* tests).

5.1 Mantel Test

The Mantel test was not significant for any of the variables within the spoken or written texts. This means that the Euclidean syntactic distances – i.e. the difference
in adverb placement between pairs of locations, for any of the four adverb classes was not a significant correlate for either as-the-crow-flies distance or TLGI.

5.2 Global Moran’s I and Local Getis-Ord Gi*

None of the variables within the written data was found to display significant spatial patterning. Therefore, in what follows, I report the findings for spoken data.

The map in Figure 5.1 displays the relative frequencies of frequency/temporal adverbs in preverbal position across the ten locations. Global Moran’s I is found to be near-significant and exhibits positive spatial autocorrelation \((I = 0.36, p = 0.018)\). In other words, frequency/temporal adverbs in preverbal position demonstrated near-significant spatial clustering overall. Figure 5.2 displays the local Getis-Ord Gi* values across the ten locations for frequency/temporal adverbs in sentence-initial position, and none of the locations exhibited significant spatial clustering of low or high values.

In Figure 5.3 and Figure 5.4, the relative frequencies and local Getis-Ord Gi* values, respectively, for postverbal frequency/temporal adverbs are mapped out. Moran’s I is found to be significant and exhibits positive spatial autocorrelation \((I = 0.66, p < 0.01)\). For varieties in (South-)East Asia (i.e., India, Singapore, Hong Kong, and the Philippines) and in East Africa (i.e., Kenya and Tanzania), they tend to pattern in a similar way in that their Gi* values are all negative. In contrast, the Gi* values for varieties in Canada, Jamaica, Northern Ireland, and the Republic of Ireland are all positive. Moreover, the Gi* scores show that Northern Ireland and the Republic of Ireland varieties exhibited significant spatial clustering of high values. This means that Ireland acted as a hot spot for placing frequency/temporal adverbs in postverbal position, and speakers of English in locations nearby Ireland tend to also place adverbs after the verb.
Figure 5.5 displays the relative frequencies of preverbal manner adverbs. Moran’s I is found to be significant and exhibits positive spatial autocorrelation ($I = 0.71, p < 0.01$). Figure 5.6 displays the local Getis-Ord Gi* values for preverbal manner adverbs with Canada, Northern Ireland and the Republic of Ireland exhibiting significant spatial clustering of low values, and India exhibiting significant spatial clustering of high values. Note that English varieties in Canada, Jamaica, and Ireland all have negative values, whereas the (South-)East Asian and East African English varieties have positive values.

Figure 5.7 displays the relative frequencies of manner adverbs in postverbal position. Moran’s I is found to be significant and exhibits positive spatial autocorrelation ($I = 0.63, p < 0.01$). The map in Figure 5.8 displays the local Getis-Ord Gi* values for manner adverbs in postverbal position with Northern Ireland and the Republic of Ireland exhibiting significant spatial clustering of high values. Again, the Canadian, Jamaican and Northern/Southern Irish English varieties pattern in a geographically similar way, as do the (South-)East Asian and East African varieties. However, the Gi* scores are opposite to what we observed in Figure 5.6, which also displays results for manner adverbs but in preverbal position.

Figure 5.9 displays the relative frequencies of modal adverbs in sentence-initial position. Moran’s I is found to be significant and exhibits positive spatial autocorrelation ($I = 0.53, p < 0.01$). Finally, the map in Figure 5.10 displays the local Getis-Ord Gi* values for modal adverbs in sentence-initial position, and none of the locations exhibited significant spatial clustering of low or high values.

Figure 5.11 displays the relative frequencies for postverbal manner adverbs. Moran’s I is found to be significant and exhibits positive spatial autocorrelation ($I = 0.54, p < 0.01$). The map in Figure 5.12 displays the local Getis-Ord Gi* values for postverbal manner adverbs with Singapore, Hong Kong and the Philippines.
exhibiting significant spatial clustering of low values. Spatially, the English varieties in (South-)East Asia pattern similarly, as do the varieties in Canada, Jamaica, and Ireland. While varieties in Kenya and Tanzania have comparable placement preferences to the Asian varieties in many of the other significant variables, the Kenyan and Tanzanian Gi* values are close to zero. Therefore, it is difficult to tell if one should consider the placement preferences of the East African Englishes to be geographically alike to the Asian Englishes, or to the other varieties.

Figure 5.1: Relative frequencies in frequency/temporal adverbs in preverbal position.
Figure 5.2: Local Getis-Ord Gi* values for frequency/temporal adverbs in preverbal position.

Figure 5.3: Relative frequencies for frequency/temporal adverbs in postverbal position.
Figure 5.4: Local Getis-Ord Gi* values for frequency/temporal adverbs in postverbal position.

Figure 5.5: Relative frequencies for manner adverbs in preverbal position.
Figure 5.6: Local Getis-Ord Gi* values for manner adverbs in preverbal position.

Figure 5.7: Relative frequencies for manner adverbs in postverbal position.
Figure 5.8: Local Getis-Ord Gi* values for manner adverbs in postverbal position.

Figure 5.9: Relative frequencies for modal adverbs in sentence-initial position.
Figure 5.10: Local Getis-Ord Gi* values for modal adverbs in sentence-initial position.

Figure 5.11: Relative frequencies for modal adverbs in preverbal position.
Figure 5.12: Local Getis-Ord Gi* values for modal adverbs in preverbal position.
Chapter 6

Discussion

The analysis of global spatial autocorrelation identified frequency/temporal adverbs in pre- and postverbal positions, pre- and postverbal manner adverbs, and modal adverbs in sentence-initial and postverbal position as exhibiting (near-)significant overall spatial clustering within the spoken data. In contrast, no variables were found to exhibit any significant geographical patterning within the written data, which indicate that adverb placement exhibits a weaker geolinguistic signal in the written register. These results show that geographical patterning exists for the ten World Englishes analyzed in this study, and demonstrate that syntactic variation, at least for variation in adverb position, in geographically and socially non-contiguous varieties correlates with geographical distance. A distance cutoff of 8,550km also found similar results to a cutoff of 6,000km for global spatial autocorrelation.

It is worth mentioning that while modal adverbs in sentence-initial and preverbal positions were found to display significant clustering, one must be wary that one position is not opposite to the other. In other words, if we know that a speaker did not place a modal adverb in preverbal position, we cannot conclude that the speaker has placed it in sentence-initial position because they could very well have placed it in postverbal position. On the other hand, preverbal manner adverbs do seem to be in complementary distribution to postverbal manner adverbs because manner adverbs almost categorically do not appear sentence-initially (see Figure 4.1).
In observing the local Getis-Ord Gi* scores across the significant variables, two recurring trends emerge. First, Canadian, Jamaican, and Northern/Southern Irish English tend to pattern in a similar way. For example, in the spoken data, Ireland (i.e. Northern Ireland and the Republic of Ireland) was identified as a hot spot for low values for manner adverbs in preverbal position and a hot spot for high values in postverbal position. Moreover, Canada and Jamaica share similar z-scores with Ireland, and this pattern was consistent among the other significant variables. The second trend is that colonial varieties in East Africa (i.e. Kenya and Tanzania) and in (South-)East Asia (i.e. India, Hong Kong, Singapore, and the Philippines) also tend to pattern in a similar way.

These results suggest that Canadian, Jamaican and the Irish Englishes form a single cluster, and the East African and (South-)East Asian Englishes form another group. These categorizations are similar to Kachru’s (1985, 1992) Three Circle model of English: the Inner Circle (e.g. English-as-a-Native-Language (ENL) varieties), the Outer Circle (e.g. English-as-a-Second-Language (ESL) varieties), and the Expanding Circle (e.g. English-as-a-Foreign-Language (EFL) varieties). Since none of the varieties analyzed in this study are considered EFL varieties, only the former two are of interest here. Several scholars have criticized that the boundaries of these circles are somewhat fuzzy (see Jenkins 2015, and Galloway and Rose 2015 for discussions regarding the criticisms of this tripartite model), although Kachru (as cited in Bolton 2006) was already aware of these gray areas:

“The Outer Circle and the Expanding Circle cannot be viewed as clearly demarcated from each other; they have several shared characteristics, and the status of English in the language policies of such countries changes from time to time. What is an ESL region at one time may become an EFL region at another time or vice versa.”
Nevertheless, the typological distinctions of the varieties based on these ENL vs. ESL (or native vs. non-native) appear to be a strong indicator of how speakers of these ten English varieties tend to pattern.\footnote{It is important to mention that while these groupings may be interesting to some people, the aim here is not to provide a characterization of the ten English varieties analyzed in this study. As have been noted by several scholars (Nerbonne 2009; Szmrecsanyi 2008, 2011, 2014), a more holistic approach whereby multiple features are aggregated would provide a more accurate and robust characterization than an individual-based approach.}

The recent findings of Szmrecsanyi et al. (2016) provides further support for the similarities found in these native and non-native groupings. In their study, they analyzed three syntactic alternations – genitive, dative, and particle placement – in four English varieties using data from the ICE corpus. They found that ENL varieties tend to pattern in a similar, as do ESL varieties. Furthermore, in their paper presented at the 2016 LSA Annual Meeting held in Washington, D.C., they extended their analysis to nine varieties of English, and in contrast, observed that the ENL and ESL varieties did not consistently pattern in a similar way as was previously suggested. Therefore, the patterning of the speakers’ syntactic choice in variant along these L1 and L2 varieties might not be explained in whole by the ‘linguistic ecology’ (or proficiency of the speakers), but by other factors.

Critics of the Kachru’s Three Circles of English have indicated that the model does not take into account the bi- and multilingualism that exist in many of these societies (e.g. Bruthiaux 2003). For example, Canada recognizes English and French as official languages, and in addition to Jamaican English, Jamaican Creole forms an important part of Jamaican identity (Schneider 2007). Therefore, it is important to consider not only the proficiencies of the speakers, but also the possibility of contact-induced changes that may be driven by factors such as intense contact and multilingualism.
within these communities. In fact, research in second-language acquisition have suggested that adverb placement is difficult to acquire even for advanced learners (White 1991).² In these studies, they find that L1 language transfer effects play a prominent role in the acquisition of adverb placement, which is reflected in the linguistic knowledge of the learners through questionnaires and preference/judgment tasks. As such, second-language acquisition strategies may be important mechanisms by which these placement tendencies emerge.

In what follows, I discuss possible explanations of this variation, focusing on the results of the spoken data, and outline several directions for future research. I also make no distinction between sentence-initial and preverbal position for convenience.

6.1 Effects of Contact Languages

In this section, I examine adverb placement in the contact languages of several locales by exploring grammars and literature discussing their possible positions. Doing so may shed light on placement preferences among the ten varieties of English. If adverb placement in the contact language(s) parallels with the distribution of adverb placement in English, then contact may have a significant influence on the speakers’ use of English. For example, in Irish, which has affected the formation of English in Ireland, adverbs and adverbials appear postverbally (McCloskey 2011). Irish may therefore have influenced the preference for postverbal position among English speakers in Ireland which was identified as a hot spot of high values for adverbs in postverbal position. In India, which exhibits spatial clustering of high values for adverbs in preverbal position, multilingualism is ubiquitous within the society. Most of the official

²See Ayoun 2005 who suggests that learners of English at the advanced stages (e.g. high school and university level) are in fact able to successfully acquire adverb placement.
languages of India (i.e. those part of the so-called Eighth Schedule to the Constitution) are in the Dravidian, Indo-Aryan, and Tibeto-Burman language families that typically have verb-final word order. For instance, in Hindi, which is the most widely spoken language in India, adverbs usually appear before the verb (Sharma 1975). Furthermore, most people in India speak English as their second language, and these speakers would be prone to interference such as syntactic transfer effects from their first language (Thomason 2001).

As we explore the languages of the other locations, there appears to be a fair number of inconsistencies between how English speakers regularly place adverbs and how adverbs are actually placed in the contact languages of the locale. While speakers in Canada have a strong tendency to place adverbs in English postverbally, similar to English speakers in Ireland, this preference is not categorical in French, in which its speakers may position them preverbally or postverbally depending on the construction.\(^3\) In simple tense clauses, adverbs must be placed postverbally:

(25) Hawkins and Towell (2015: 132, emphasis theirs)
   a. Je veux **souvent** le faire.
   b. * Je **souvent** veux le faire.

Moreover, in the so-called compound tense clauses, adverbs may appear in several positions within a sentence:

(26) Hawkins and Towell (2015: 132, emphasis theirs)
   a. J’ai **souvent** voulu le faire.
   b. J’ai voulu **souvent** le faire.

\(^3\)According to the manual of the ICE-Canada corpus (Newman and Columbus 2010), most of the data were collected from speakers living in Ontario and Quebec. While the majority of the informants spoke English as their first language, a great number also spoke French.
According to Hawkins and Towell (2015), the placement of adverbs is influenced by the number of syllables of the adverb – monosyllabic words typically appear clause-internally, and polysyllabic words usually appear clause-initially or -finally. Abeillé and Godard (2004) even notes that “there are adverbs which are always postverbal,” even in compound tense clauses (e.g. *verticalement* ‘vertically’, *horizontalement* ‘horizontally’, *spacieusement* ‘spaciously’, *mortellement* ‘mortally’).

In sum, the preference for preverbal and postverbal position among speakers of English in India and Ireland, respectively, parallels with the placement of adverbs in the contact languages. In Hindi, which is spoken by the majority in India, adverbs appear before the verb. In Irish, adverbs appear after the verb. However, in French, where there is a large speaker population in Canada, adverbs may appear before or after the verb, yet speakers of Canadian English typically place adverbs after the verb. It is important to note that grammars only represent a synchronic view of the language. Furthermore, due to the lack of diachronic corpora in general (and even more so for underrepresented languages) from which adverb placement frequencies can be extracted, it may be challenging to accurately explain the effects of the contact languages which would have been influencing the Englishes from their inception onwards.

Nevertheless, if the effect of contact is found to be a significant mechanism of variation and change in the placement preferences of adverbs across the ten English varieties, the overarching question remains: why exactly do we find geographical patterning of adverb placement across these geographically and socially non-contiguous World Englishes? In a study by Holman et al. (2007) investigating the relationship between typological similarity and geographical distance using the World Atlas of Language Structures, they found that languages closer in proximity, even if they are not related, share more linguistic features than languages farther away. In the same
line of thinking, increased geographical distance may not actually be a direct correlate with greater differences in adverb placement among these English varieties. Instead, the relation may be between geographical distance and adverb placement in the contact languages. The contact languages are simply influencing adverb placement in these English varieties. In other words, India and Singapore are more geographically proximate than India and Canada. Because of this, we would expect local languages spoken in India and languages spoken in Singapore to be structurally more similar. The effects may be diametrically opposite for local languages spoken in India and languages spoken in Canada. As a result of intense contact, English varieties worldwide should then mirror the structures of the other languages spoken in their respective locations resulting in a similar geographical patterning reported in the study by Holman et. al.

6.2 Similarities to Historic Englishes

Trudgill and Hannah (2008) observe that Englishes of several diasporas, such as the varieties in the Southern Hemisphere, share many features based on similar sociohistorical contexts such as the periods in which colonization first occurred:

“Not surprisingly, these patterns of expansion, settlement and colonization have had an effect on the relationships, similarities and differences between the varieties of English which have grown up in different parts of the world. For example, [...] the English varieties of the Southern Hemisphere (Australia, New Zealand, South Africa, Falklands), which were transplanted relatively recently from the British Isles, are very similar to each other. They are quite naturally much less different from the English of England
than are the varieties spoken in the Americas, which were settled much earlier.”

(Trudgill and Hannah 2008: 9)

This suggests that the ten English varieties may share similar adverb placement frequencies to the Englishes that were first introduced in these locations due to colonization. Therefore, varieties that exhibit similar placement preferences may be derived from the use of the Englishes spoken around the same period. In other words, varieties of English spoken in India and Singapore, which tend to place adverbs preverbally, may have been colonized around the same time, and so speakers of English at the time of colonization may also prefer to place adverbs preverbally.

<table>
<thead>
<tr>
<th>Location</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Year</th>
<th>Colonizers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada</td>
<td>56.130366</td>
<td>-106.346771</td>
<td>1537</td>
<td>England &amp; France</td>
</tr>
<tr>
<td>Jamaica</td>
<td>18.109581</td>
<td>-77.297508</td>
<td>1655</td>
<td>England</td>
</tr>
<tr>
<td>Northern Ireland</td>
<td>54.787715</td>
<td>-6.492315</td>
<td>1536</td>
<td>England</td>
</tr>
<tr>
<td>Republic of Ireland</td>
<td>53.41291</td>
<td>-8.24389</td>
<td>1536</td>
<td>England</td>
</tr>
<tr>
<td>Tanzania</td>
<td>-6.369028</td>
<td>34.888822</td>
<td>1919</td>
<td>Great Britain</td>
</tr>
<tr>
<td>Kenya</td>
<td>-0.023559</td>
<td>37.906193</td>
<td>1895</td>
<td>Great Britain</td>
</tr>
<tr>
<td>India</td>
<td>20.593684</td>
<td>78.96288</td>
<td>1830</td>
<td>Great Britain</td>
</tr>
<tr>
<td>Singapore</td>
<td>1.352083</td>
<td>103.819836</td>
<td>1819</td>
<td>Great Britain</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>22.396428</td>
<td>114.109497</td>
<td>1842</td>
<td>Great Britain</td>
</tr>
<tr>
<td>Philippines</td>
<td>12.879721</td>
<td>121.774017</td>
<td>1898</td>
<td>United States</td>
</tr>
</tbody>
</table>

Table 6.1: Year of initial colonization for each location (in ascending longitude).

The years in which the ten locations were colonized are displayed in Table 6.1. By associating the year of initial colonization of each location to the English that was first introduced, we can observe a similar trend in the placement preferences reported in the results. Tanzania, Kenya, India, Singapore, Hong Kong and the Philippines, where speakers of English tend to place adverbs preverbally, were colonized in the 19th and 20th century, whereas Canada, Jamaica and Ireland, where speakers tend to place
adverbs postverbally, were colonized pre-19th century. This suggests that speakers of English in the 19th and 20th century prefer to place adverbs before the verb, whereas speakers of English before the 19th century prefer to place adverbs after the verb. Fortunately, there are many corpora of historic Englishes due to the abundance of texts such as letters and parliament proceedings that have been preserved. Extracting adverb placement frequencies from these corpora may be able to provide support for this hypothesis. If it is found that placement preferences in the English varieties today are similar to placement preferences in the historic Englishes, then it would suggest that the geographical patterning that we observe in the data may be the result of colonization of multiple neighboring locations within the same era. By way of illustration, once the British empire colonized Singapore in 1819, it would have been more economical to colonize other locations within the vicinity than locations farther away. This would then lead to similar varieties of historic Englishes being introduced within the region, and thus, similarities in placement patterns in the modern Englishes of these locales.

6.3 Probabilistic Knowledge and Indigenization

Scholars working on syntactic variation within the so-called Probabilistic Grammar framework, postulate that (a) the choice in syntactic variant is driven by and are sensitive to probabilistic and gradient factors, and (b) linguistic knowledge is partially (or fundamentally) probabilistic. In their analysis of several syntactic alternations (genitive and dative alternation, and particle placement) across World Englishes, Szmrecsanyi et al. (2016) attribute the variation among the varieties to ‘probabilistic indigenization’, that is, “the process whereby stochastic patterns of internal linguistic variation are reshaped by shifting usage frequencies in speakers of post-colonial vari-
eties. To the extent that patterns of variation in a new variety A [...] can be shown to differ from those of the mother variety, we can say that the new pattern represents a novel, if gradient, development in the grammar of A." If their theory is correct, this would still not explain why geographical patterns arise and why speakers in the Western Hemisphere tend to place adverbs postverbally, while speakers in the Eastern Hemisphere tend to place adverbs preverbally. In other words, what is the cause for the process of probabilistic indigenization or nativization among these speakers? While this is not explored in this thesis, this is an area for fruitful research that may be able to contribute towards the theory that speakers' syntactic knowledge is partially probabilistic.4

6.4 Diachronic Syntax

Language variation is intimately connected with language change, and for many scholars, it is also linked to language acquisition and learnability (Gibson and Wexler 1994, Fodor 1998, Dresher 1999, Lightfoot 1999, Roberts 2007). Changes occur when the input in one generation differs from the previous generation leading to differences

4One possible predictor of adverb placement may involve the principle of end weight (Wasow 2002), that is, if a clause contains a long and complex phrase, this phrase will tend to be placed at the end of the clause. Therefore, in these constructions, adverbs will tend to appear preverbally where it is closer to the verb it is modifying, than postverbally, that is, after the long and complex phrase. In a preliminary study, I found that adverbs tend to be placed in preverbal position as the length of the verbal object, measured by the number of tokens, increases.

(27) a. We now recycle glass and newspapers and aluminum cans and plastics.
    b. % We recycle glass and newspapers and aluminum cans and plastics now.

This suggests that given two alternating constructions in (27), the sentence where the adverb is placed preverbally, which was actually extracted from the Switchboard corpus, a corpus of US English telephone conversations (Godfrey et al. 1992), is preferred over postverbal position, indicated by the percent symbol. However, further investigation is needed to support this claim.
in the internal grammar of both generations (Lightfoot 2006). In other words, changes in the input (or primary linguistic data) trigger new grammars being acquired, and these new grammars then produce novel input for the next generation. Under this view, variation happens due to differences in input across several communities through several generations.

If we assume that differences in the input is the source of variation in the placement preferences of adverbs, then what might have caused this shift? One might propose second-language learning strategies, transfer effects from second-language acquisition, multilingualism and other factors that affect the use of English across these communities to play a role, just as it has been hypothesized that the morphological simplification in the history of English is due to widespread English-Scandinavian bilingualism (O’Neil 1978).

Much of the work on historical (or diachronic) syntax focuses on monolingual first language acquisition, but bi-/multilingual first and second-language acquisition are ubiquitous, and provide important insights for understanding how and why certain changes occur. Furthermore, it would also be helpful to incorporate these ideas into theories such as coexisting grammars (Kroch 1989, 1991), which have been employed to explain variation within a speech community or within an individual, in order to achieve higher levels of explanatory adequacy. According to Kroch’s theory, the community or the individual oscillates between two (or more) competing grammars with one grammar replacing the others at a constant rate (also known as the Constant Rate Hypothesis). More recently, the findings of Han et al. (2007) suggest that individuals acquire only one grammar even if there is variation within a speech community. They reported individual differences in scope judgments of the interaction between QP and negation in Korean, where half the speakers allowed only the interpretation for the universal to take scope over the negation, and the other half allowed only negation...
to take scope over the universal. They argue that “learners choose one grammar at random and discard the other option.” However, as Roberts (2007) notes, the “possibility that a single individual may have two distinct grammars must of course be acknowledged for the case of true bilinguals.” But where does one draw the line between two grammars being acquired successfully (e.g. Chinese-English bilinguals) and one grammar chosen at random while the other is tossed aside (e.g. scope over quantifier or negation in Korean)? I do not attempt to resolve these issues here, but leave them for future considerations.

6.5 Limitations

Despite the significant findings of this study, there are several important limitations I would like to point out. The first limitation is the issue of semantic equivalence. As discussed in Section 4.4, there are subtle nuances in the placement of aspect-manner (e.g. quickly), resultative (e.g. beautifully), frequency (e.g. frequently), and even temporal (e.g. recently) adverbs, and it was not possible to analyze pure manner adverbs due to the very low counts in the texts.\(^5\)

There are several reasons why this should not significantly affect the results of this study. The first reason is that many of the English varieties analyzed in this study are L2 varieties (e.g. Singapore, Hong Kong, India, Philippines) – i.e. many speakers of these (South-)East Asian population centers speak English as their second language. Many studies have shown that L1 syntactic transfer effects play a prominent role in the placement of adverbs for L2 speakers of English (White 1991, Trahey 1996, Chan 2004, Chu and Schwartz 2005, among others). For example, White (1991) finds that French

\(^5\)It is worth pointing out that the Corpus of Web-Based Global English (GloWbE) contains 1.9 billion words of twenty varieties of English worldwide, and so there may be enough pure manner adverbs, which have been argued to be truly semantically equivalent (Ernst 2002), to control for meaning.
learners of English frequently place adverbs between the verb and the complement (e.g. *Logan ate often the pizza). She suggests that this incorrect placement is due to their L1 (i.e. French), which allows adverbs to appear between the verb and the complement (e.g. *Pierre mange souvent la baguette). This suggests that L2 speakers of English are influenced by well-formed sentence structures to a higher degree than the subtle differences in meaning.

Secondly, among bilinguals and multilinguals, languages are active and competing. A Chinese-English bilingual conversing in Chinese is accessing their grammar of Chinese. At the same time, their English grammar remains active rather than being dormant until it is required. When a bilingual hears a speech signal, it has been suggested that they attempt to identify lexical items within the signal by activating both grammars in parallel in varying degrees (Marian and Spivey 2003). Moreover, while bi- and multilinguals are generally considered to have separate autonomous grammars, there is a possibility for cross-linguistic influence due to the fact that an individual, who knows more than one language, shares cognitive resources for pragmatic and discourse processes. Because of this, it has been noted that the coordination of syntax and pragmatics across multiple languages is a difficult task in bi- and multilingual first language acquisition (Serratrice et al. 2004). This differs from second-language learners who rely heavily on their first language during the learning process. In their study, Valian and Eisenberg (1996) find that bilingual children acquiring a non-null subject language (e.g. English) as well as a null subject language (e.g. Italian) omit subjects significantly more often than monolinguals of a non-null subject language. In other words, an English-Italian bilingual child is predicted to omit subjects more often than an English monolingual child. This suggests that children are aware of the frequencies of null-subjects in their language. Thus, if the choice of adverb placement is conditioned in part by pragmatic constraints, then we would expect the effects of
bi- and multilingual first language acquisition at the syntax-pragmatics interface to play a role in the placement of adverbs regardless of the slight meaning difference.

The third reason is that the nuances in meaning of several adverbs are incredibly subtle and while it has been universally agreed that adverbs can differ in meaning, there is still no general consensus as to which adverbs may remain semantically equivalent. For example, Jackendoff (1972: 42) lists *quickly*, *slowly* and *frequently* as those that have no “discernible change in meaning”, whereas *quickly* and *slowly* are what Ernst (2002) classifies as aspect-manner adverbs which have subtle differences in their meanings as does the frequency adverb, *frequently*. Moreover, *tightly* is among those that Ernst (2002) calls pure manner adverbs, which do not alter in meaning. However, it can be argued that *tightly* functions in a similar way to resultative adverbs:

(28) a. Jordan *tightly* sealed the lid.
    
    b. Jordan sealed the lid *tightly*.

In (28a), there is only a manner reading in which Jordan sealed the lid in a tight way, but in (28b), it is ambiguous between a manner and resultative interpretation whereby the latter signifies the result of the action, and the resultant object is a tight lid. Therefore these inconsistent accounts supported by decontextualized examples raise questions as to what exactly are the differences in meaning, if any. The question of semantic equivalence is not unique to adverbs, but has also been widely debated in other well-known syntactic alternations (see Bresnan and Ford 2010 for discussion on the semantics of dative shift). Furthermore, others have even postulated that variants may become neutralized in discourse (Sankoff 1989), but even this hypothesis leaves much to be explained.⁶

---

⁶In his response to Beatriz Lavandera's critiques on the viability of syntactic variants as a linguistic variable, Labov (1978: 5) states the following, “But we are not in the business of being persuasive: our enterprise demands conclusive demonstration.”
Lastly, there are several studies where meaning differences were not controlled for, but still, some of their findings have made important contributions. The study by White (1991) on the acquisition of adverb placement had been seminal work in investigating the issue of parameter acquisition within the generative paradigm. Adverb placement has been used as a diagnostic for verb-movement, and it has been argued that English is a non-verb-movement language, whereas French is (Pollock 1989). However, in her study, White (1991: 141) notes that “there was no attempt to teach the kind of subtle meaning differences that can result from placing adverbs in the various different positions.” In another study, Grieve (2012b) analyzed adverb position across written records of standard American English. Specifically, he looked at ‘common’ adverbs, and adverbs longer than five characters that end in -ly, which would consist of adverbs that change in meaning. Nevertheless, many of the variables in the study consistently identified the dialects of the Northeast against those in the Southeast and South Central states.

The second limitation concerns the sample size. Naturally, the small sample size (i.e. the small number of locations) means that the power of the results is considerably low. Unsurprisingly, one way to increase the reliability of these tests is to increase the number of locations, or add more varieties of English into the repertoire. While there are many existing corpora of North American English (e.g. Switchboard Corpus), this variety was not included simply because these corpora do not follow the same common design employed by the ICE corpus. The GloWbe Corpus, which is a 1.9 billion word corpus featuring twenty World English varieties, is a potential next step in order to address the issues that follow from a limited sample size, and to determine whether these results generalize to other World Englishes. It is also not clear to me why the Mantel test statistic did not find any of the variables to be significant whereas
global Moran’s I and local Getis-Ord Gi* found five variables to display significant clustering, but I suspect the low sample size may play a role.

Finally, the corpus design may also pose problems. Apart from Kenya, Tanzania, Northern Ireland, and the Republic of Ireland, which consist of 500,000 words each, the other corpora contain one million words. These corpora are further divided into their spoken and written registers. This means that the East African and Irish varieties have half the number of words in both spoken and written texts in comparison to the other corpora. This can be an issue because the distribution of adverbs may differ with corpus size. Furthermore, these corpora were compiled during different periods and while these corpora follow a ‘common’ design, it is up to the compilers how they interpret these guidelines. Thus, this introduces temporal and genre variation into the data.
Chapter 7

Conclusion

This thesis has shown that adverb placement exhibits significant spatial patterning among ten (non-)contiguous varieties of World Englishes in the spoken register. Frequency/temporal, manner and modal adverbs were found to display significant global patterning. The (South-)East Asian varieties were identified as major hot spots for placing adverbs before the verb, whereas Canada and Ireland acted as the main hot spots for post-verbal placement. None of the variables within the written texts were found to display any spatial patterning, suggesting a much weaker geolinguistic signal in the written texts in comparison to the spoken register. This also supports Schneider’s (2002) claim that we should not find any strong regional or sociolinguistic variation within written documents because they conceal much of an author’s speech patterns.

Furthermore, the geographical patterning shares many similarities to Kachru’s (1985, 1992) Three Circles of English model, which groups World Englishes according to the ‘average’ speaker’s level of proficiency (i.e. native vs. non-native). Because many people in these societies speak a language other than English, we might expect the effects of contact, and bi- and multilingualism to play a role in the placement preferences in these locales. This is supported by findings in the second-language acquisition literature which reports that adverb placement in English is difficult to acquire for L2 learners and suggest that syntactic transfer effects from the L1 are a significant factor (White 1991, Trahey 1996, Chu and Schwartz 2005, among others).
While these results are important for identifying spatial clustering among English varieties worldwide, they do not provide a full answer to why these patterns exist in the first place and whether this is actual ongoing change across the English varieties. This thesis takes initial steps toward answering these questions, with the hope that future investigations may want to explore the effects of contact languages and historic Englishes, among other potential factors. The Corpus of Web-Based Global English (GloWbE), which contains texts from twenty World Englishes, is another set of data that could be analyzed in order to increase the reliability of the results due to the low number of locations investigated in this study, and to provide further evidence for the significance of the spatial patterning. Because the GloWbE corpus contains around 1.9 billion words, one can restrict the class of manner adverbs to just pure manner adverbs (e.g. loudly).

Finally, this thesis contributes to the growing body of research in geolinguistics, and complements existing works on the relationship between geography and syntactic variation through the use of naturalistic corpora and statistical techniques. Language variation is intimately linked to language change and so geostatistical techniques that are able to reveal underlying spatial patterns may also reflect the outcomes of language change. Therefore, these findings raise various questions regarding the source of variation in adverb position among World Englishes. They also open up an abundance of research avenues into variation and change across World Englishes from a variety of perspectives.
Appendix A

Appendix: List of Adverbs

**Evaluative adverbs:** hopefully, surely, importantly, fortunately, significantly, supposedly, presumably, allegedly, luckily, inevitably, undoubtedly, admittedly, somehow, thankfully, evidently

**Frequency/temporal adverbs:** regularly, usually, finally, normally, recently, immediately, eventually, suddenly, frequently, currently, ultimately, previously, gradually, shortly, constantly, lately, rarely, occasionally, simultaneously, continuously, consistently, temporarily, annually, repeatedly, continually, instantly, immediately, again, sometimes, momentarily

**Manner adverbs:** quickly, slowly, closely, heavily, rapidly, quietly, smoothly, beautifully, comfortably, lightly, emotionally, legally, loudly, sharply, voluntarily, gently, purposely, tightly, illegally, violently, neatly, patiently, softly, fiercely, loosely, hurriedly, convincingly, brightly, musically, cleanly, fatally, passionately, powerfully, gracefully

**Modal adverbs:** probably, probly, certainly, definitely, necessarily, possibly, unlikely, likely, unnecessarily, maybe, perhaps
## Appendix B

### Appendix: Adverb Placement Frequencies from the Spoken Texts

<table>
<thead>
<tr>
<th>Adverb</th>
<th>Country</th>
<th>Initial</th>
<th>Preverbal</th>
<th>Postverbal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluator</td>
<td>Canada</td>
<td>27</td>
<td>42.9%</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>Jamaica</td>
<td>71</td>
<td>46.4%</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>Northern Ireland</td>
<td>49</td>
<td>55.7%</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>Republic of Ireland</td>
<td>30</td>
<td>35.3%</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>Tanzania</td>
<td>9</td>
<td>29.0%</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Kenya</td>
<td>27</td>
<td>40.3%</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>India</td>
<td>41</td>
<td>37.3%</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>Singapore</td>
<td>78</td>
<td>47.6%</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Hong Kong</td>
<td>76</td>
<td>47.5%</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>Philippines</td>
<td>75</td>
<td>40.1%</td>
<td>71</td>
</tr>
</tbody>
</table>

Table B.1: Placement counts and percentages for evaluative adverbs in the spoken texts.

<table>
<thead>
<tr>
<th>Adverb</th>
<th>Country</th>
<th>Initial</th>
<th>Preverbal</th>
<th>Postverbal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>Canada</td>
<td>229</td>
<td>26.7%</td>
<td>337</td>
</tr>
<tr>
<td></td>
<td>Jamaica</td>
<td>224</td>
<td>28.5%</td>
<td>298</td>
</tr>
<tr>
<td></td>
<td>Northern Ireland</td>
<td>115</td>
<td>24.7%</td>
<td>179</td>
</tr>
<tr>
<td></td>
<td>Republic of Ireland</td>
<td>115</td>
<td>26.1%</td>
<td>153</td>
</tr>
<tr>
<td></td>
<td>Tanzania</td>
<td>200</td>
<td>28.4%</td>
<td>280</td>
</tr>
<tr>
<td></td>
<td>Kenya</td>
<td>205</td>
<td>28.5%</td>
<td>280</td>
</tr>
<tr>
<td></td>
<td>India</td>
<td>243</td>
<td>30.4%</td>
<td>275</td>
</tr>
<tr>
<td></td>
<td>Singapore</td>
<td>246</td>
<td>26.1%</td>
<td>353</td>
</tr>
<tr>
<td></td>
<td>Hong Kong</td>
<td>335</td>
<td>30.1%</td>
<td>394</td>
</tr>
<tr>
<td></td>
<td>Philippines</td>
<td>221</td>
<td>25.5%</td>
<td>363</td>
</tr>
</tbody>
</table>

Table B.2: Placement counts and percentages for frequency/temporal adverbs in the spoken texts.
<table>
<thead>
<tr>
<th>Adverb</th>
<th>Country</th>
<th>Initial</th>
<th>Preverbal</th>
<th>Postverbal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manner</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canada</td>
<td>3</td>
<td>1.5%</td>
<td>97</td>
<td>47.1%</td>
</tr>
<tr>
<td>Jamaica</td>
<td>5</td>
<td>2.4%</td>
<td>82</td>
<td>39.6%</td>
</tr>
<tr>
<td>Northern Ireland</td>
<td>1</td>
<td>0.9%</td>
<td>49</td>
<td>42.2%</td>
</tr>
<tr>
<td>Republic of Ireland</td>
<td>3</td>
<td>3.8%</td>
<td>39</td>
<td>48.8%</td>
</tr>
<tr>
<td>Tanzania</td>
<td>0</td>
<td>0.0%</td>
<td>89</td>
<td>47.1%</td>
</tr>
<tr>
<td>Kenya</td>
<td>3</td>
<td>1.8%</td>
<td>77</td>
<td>47.0%</td>
</tr>
<tr>
<td>India</td>
<td>4</td>
<td>2.2%</td>
<td>90</td>
<td>49.2%</td>
</tr>
<tr>
<td>Singapore</td>
<td>13</td>
<td>4.9%</td>
<td>96</td>
<td>43.0%</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>4</td>
<td>1.7%</td>
<td>103</td>
<td>43.3%</td>
</tr>
<tr>
<td>Philippines</td>
<td>10</td>
<td>4.7%</td>
<td>95</td>
<td>44.4%</td>
</tr>
</tbody>
</table>

Table B.3: Placement counts and percentages for manner adverbs in the spoken texts.

<table>
<thead>
<tr>
<th>Adverb</th>
<th>Country</th>
<th>Initial</th>
<th>Preverbal</th>
<th>Postverbal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canada</td>
<td>183</td>
<td>40.0%</td>
<td>156</td>
<td>34.1%</td>
</tr>
<tr>
<td>Jamaica</td>
<td>118</td>
<td>39.9%</td>
<td>106</td>
<td>35.8%</td>
</tr>
<tr>
<td>Northern Ireland</td>
<td>85</td>
<td>31.0%</td>
<td>104</td>
<td>38.0%</td>
</tr>
<tr>
<td>Republic of Ireland</td>
<td>49</td>
<td>23.0%</td>
<td>78</td>
<td>36.6%</td>
</tr>
<tr>
<td>Tanzania</td>
<td>78</td>
<td>27.4%</td>
<td>90</td>
<td>31.6%</td>
</tr>
<tr>
<td>Kenya</td>
<td>73</td>
<td>28.0%</td>
<td>87</td>
<td>33.3%</td>
</tr>
<tr>
<td>India</td>
<td>104</td>
<td>33.4%</td>
<td>100</td>
<td>32.2%</td>
</tr>
<tr>
<td>Singapore</td>
<td>150</td>
<td>36.0%</td>
<td>144</td>
<td>34.5%</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>122</td>
<td>31.4%</td>
<td>148</td>
<td>38.1%</td>
</tr>
<tr>
<td>Philippines</td>
<td>151</td>
<td>43.4%</td>
<td>120</td>
<td>34.5%</td>
</tr>
</tbody>
</table>

Table B.4: Placement counts and percentages for modal adverbs in the spoken texts.
### Appendix C

**Appendix: Adverb Placement Frequencies from the Written Texts**

<table>
<thead>
<tr>
<th>Adverb</th>
<th>Country</th>
<th>Initial</th>
<th>Preverbal</th>
<th>Postverbal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluator</td>
<td></td>
<td></td>
<td></td>
<td>Raw</td>
</tr>
<tr>
<td></td>
<td>Canada</td>
<td>51</td>
<td>36.2%</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Jamaica</td>
<td>46</td>
<td>40.7%</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>Northern Ireland</td>
<td>28</td>
<td>39.4%</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Republic of Ireland</td>
<td>17</td>
<td>29.3%</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Tanzania</td>
<td>37</td>
<td>39.4%</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>Kenya</td>
<td>45</td>
<td>43.7%</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>India</td>
<td>42</td>
<td>37.5%</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>Singapore</td>
<td>63</td>
<td>53.4%</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>Hong Kong</td>
<td>60</td>
<td>46.2%</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>Philippines</td>
<td>52</td>
<td>42.3%</td>
<td>46</td>
</tr>
</tbody>
</table>

Table C.1: Placement counts and percentages for evaluative adverbs in the written texts.

<table>
<thead>
<tr>
<th>Adverb</th>
<th>Country</th>
<th>Initial</th>
<th>Preverbal</th>
<th>Postverbal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td></td>
<td></td>
<td></td>
<td>Raw</td>
</tr>
<tr>
<td></td>
<td>Canada</td>
<td>229</td>
<td>26.7%</td>
<td>337</td>
</tr>
<tr>
<td></td>
<td>Jamaica</td>
<td>224</td>
<td>28.5%</td>
<td>298</td>
</tr>
<tr>
<td></td>
<td>Northern Ireland</td>
<td>115</td>
<td>24.7%</td>
<td>179</td>
</tr>
<tr>
<td></td>
<td>Republic of Ireland</td>
<td>115</td>
<td>26.1%</td>
<td>153</td>
</tr>
<tr>
<td></td>
<td>Tanzania</td>
<td>200</td>
<td>28.4%</td>
<td>280</td>
</tr>
<tr>
<td></td>
<td>Kenya</td>
<td>205</td>
<td>28.5%</td>
<td>280</td>
</tr>
<tr>
<td></td>
<td>India</td>
<td>243</td>
<td>30.4%</td>
<td>275</td>
</tr>
<tr>
<td></td>
<td>Singapore</td>
<td>246</td>
<td>26.1%</td>
<td>353</td>
</tr>
<tr>
<td></td>
<td>Hong Kong</td>
<td>335</td>
<td>30.1%</td>
<td>394</td>
</tr>
<tr>
<td></td>
<td>Philippines</td>
<td>221</td>
<td>25.5%</td>
<td>363</td>
</tr>
</tbody>
</table>

Table C.2: Placement counts and percentages for frequency/temporal adverbs in the written texts.
<table>
<thead>
<tr>
<th>Adverb</th>
<th>Country</th>
<th>Initial</th>
<th></th>
<th>Preverbal</th>
<th></th>
<th>Postverbal</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Manner</td>
<td>Canada</td>
<td>3</td>
<td>1.5%</td>
<td>97</td>
<td>47.1%</td>
<td>105</td>
<td>51.0%</td>
</tr>
<tr>
<td></td>
<td>Jamaica</td>
<td>5</td>
<td>2.4%</td>
<td>82</td>
<td>39.6%</td>
<td>120</td>
<td>58.0%</td>
</tr>
<tr>
<td></td>
<td>Northern</td>
<td>1</td>
<td>0.9%</td>
<td>49</td>
<td>42.2%</td>
<td>64</td>
<td>55.2%</td>
</tr>
<tr>
<td></td>
<td>Ireland</td>
<td>3</td>
<td>3.8%</td>
<td>39</td>
<td>48.8%</td>
<td>36</td>
<td>45.0%</td>
</tr>
<tr>
<td></td>
<td>Tanzania</td>
<td>0</td>
<td>0.0%</td>
<td>89</td>
<td>47.1%</td>
<td>100</td>
<td>52.9%</td>
</tr>
<tr>
<td></td>
<td>Kenya</td>
<td>3</td>
<td>1.8%</td>
<td>77</td>
<td>47.0%</td>
<td>84</td>
<td>51.2%</td>
</tr>
<tr>
<td></td>
<td>India</td>
<td>4</td>
<td>2.2%</td>
<td>90</td>
<td>49.2%</td>
<td>88</td>
<td>48.6%</td>
</tr>
<tr>
<td></td>
<td>Singapore</td>
<td>13</td>
<td>4.9%</td>
<td>96</td>
<td>36.0%</td>
<td>154</td>
<td>57.7%</td>
</tr>
<tr>
<td></td>
<td>Hong Kong</td>
<td>4</td>
<td>1.7%</td>
<td>103</td>
<td>43.3%</td>
<td>130</td>
<td>51.2%</td>
</tr>
<tr>
<td></td>
<td>Philippines</td>
<td>10</td>
<td>4.7%</td>
<td>95</td>
<td>44.4%</td>
<td>105</td>
<td>49.1%</td>
</tr>
</tbody>
</table>

Table C.3: Placement counts and percentages for manner adverbs in the written texts.

<table>
<thead>
<tr>
<th>Adverb</th>
<th>Country</th>
<th>Initial</th>
<th></th>
<th>Preverbal</th>
<th></th>
<th>Postverbal</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Modal</td>
<td>Canada</td>
<td>183</td>
<td>40.0%</td>
<td>156</td>
<td>34.1%</td>
<td>116</td>
<td>25.3%</td>
</tr>
<tr>
<td></td>
<td>Jamaica</td>
<td>118</td>
<td>39.9%</td>
<td>106</td>
<td>35.8%</td>
<td>60</td>
<td>20.3%</td>
</tr>
<tr>
<td></td>
<td>Northern</td>
<td>85</td>
<td>31.0%</td>
<td>104</td>
<td>38.0%</td>
<td>81</td>
<td>29.6%</td>
</tr>
<tr>
<td></td>
<td>Ireland</td>
<td>49</td>
<td>23.0%</td>
<td>78</td>
<td>36.6%</td>
<td>80</td>
<td>37.6%</td>
</tr>
<tr>
<td></td>
<td>Tanzania</td>
<td>78</td>
<td>27.4%</td>
<td>90</td>
<td>31.6%</td>
<td>113</td>
<td>39.6%</td>
</tr>
<tr>
<td></td>
<td>Kenya</td>
<td>73</td>
<td>28.0%</td>
<td>87</td>
<td>33.3%</td>
<td>95</td>
<td>36.4%</td>
</tr>
<tr>
<td></td>
<td>India</td>
<td>104</td>
<td>33.4%</td>
<td>100</td>
<td>32.2%</td>
<td>106</td>
<td>34.1%</td>
</tr>
<tr>
<td></td>
<td>Singapore</td>
<td>150</td>
<td>36.0%</td>
<td>144</td>
<td>34.5%</td>
<td>116</td>
<td>27.8%</td>
</tr>
<tr>
<td></td>
<td>Hong Kong</td>
<td>122</td>
<td>31.4%</td>
<td>148</td>
<td>38.1%</td>
<td>113</td>
<td>29.1%</td>
</tr>
<tr>
<td></td>
<td>Philippines</td>
<td>151</td>
<td>43.4%</td>
<td>120</td>
<td>34.5%</td>
<td>75</td>
<td>21.6%</td>
</tr>
</tbody>
</table>

Table C.4: Placement counts and percentages for modal adverbs in the written texts.
Appendix D

Appendix: Results for a Distance Band of 3000km

![Map showing local Getis-Ord Gi* values for evaluative adverbs in sentence-initial position.](image)

Figure D.1: Local Getis-Ord Gi* values for evaluative adverbs in sentence-initial position.
Moran's I: −0.0741 , p = 0.421
Getis-Ord Gi* z-score
−2 −1 0 1 2
Canada
Jamaica
R. of Ireland
N. Ireland
Tanzania
Kenya
India
Singapore
HK
Philippines
Moran's I: −0.396 , p = 0.736
Getis-Ord Gi* z-score
−2 −1 0 1 2
Canada
Jamaica
R. of Ireland
N. Ireland
Tanzania
Kenya
India
Singapore
HK
Philippines

Figure D.2: Local Getis-Ord Gi* values for evaluative adverbs in preverbal position.

Figure D.3: Local Getis-Ord Gi* values for evaluative adverbs in postverbal position.
Figure D.4: Local Getis-Ord Gi* values for frequency/temporal adverbs in sentence-initial position.

Figure D.5: Local Getis-Ord Gi* values for frequency/temporal adverbs in preverbal position.
Figure D.6: Local Getis-Ord Gi* values for frequency/temporal adverbs in postverbal position.

Figure D.7: Local Getis-Ord Gi* values for manner adverbs in sentence-initial position.
Figure D.8: Local Getis-Ord Gi* values for manner adverbs in preverbal position.

Figure D.9: Local Getis-Ord Gi* values for manner adverbs in postverbal position.
Figure D.10: Local Getis-Ord Gi* values for modal adverbs in sentence-initial position.

Figure D.11: Local Getis-Ord Gi* values for modal adverbs in preverbal position.
Figure D.12: Local Getis-Ord Gi* values for modal adverbs in postverbal position.
APPENDIX E

APPENDIX: RESULTS FOR A DISTANCE BAND OF 6000KM

Figure E.1: Local Getis-Ord Gi* values for evaluative adverbs in sentence-initial position.
Figure E.2: Local Getis-Ord Gi* values for evaluative adverbs in preverbal position.

Figure E.3: Local Getis-Ord Gi* values for evaluative adverbs in postverbal position.
Figure E.4: Local Getis-Ord Gi* values for frequency/temporal adverbs in sentence-initial position.

Figure E.5: Local Getis-Ord Gi* values for frequency/temporal adverbs in preverbal position.
Figure E.6: Local Getis-Ord Gi* values for frequency/temporal adverbs in postverbal position.

Figure E.7: Local Getis-Ord Gi* values for manner adverbs in sentence-initial position.
Figure E.8: Local Getis-Ord Gi* values for manner adverbs in preverbal position.

Figure E.9: Local Getis-Ord Gi* values for manner adverbs in postverbal position.
Figure E.10: Local Getis-Ord Gi* values for modal adverbs in sentence-initial position.

Figure E.11: Local Getis-Ord Gi* values for modal adverbs in preverbal position.
Figure E.12: Local Getis-Ord Gi* values for modal adverbs in postverbal position.
Appendix F

Appendix: Results for a Distance Band of 8550 km

Figure F.1: Local Getis-Ord Gi* values for evaluative adverbs in sentence-initial position.
Figure F.2: Local Getis-Ord Gi* values for evaluative adverbs in preverbal position.

Figure F.3: Local Getis-Ord Gi* values for evaluative adverbs in postverbal position.
Figure F.4: Local Getis-Ord Gi* values for frequency/temporal adverbs in sentence-initial position.

Figure F.5: Local Getis-Ord Gi* values for frequency/temporal adverbs in preverbal position.
Figure F.6: Local Getis-Ord Gi* values for frequency/temporal adverbs in postverbal position.

Figure F.7: Local Getis-Ord Gi* values for manner adverbs in sentence-initial position.
Figure F.8: Local Getis-Ord Gi* values for manner adverbs in preverbal position.

Figure F.9: Local Getis-Ord Gi* values for manner adverbs in postverbal position.
Figure F.10: Local Getis-Ord Gi* values for modal adverbs in sentence-initial position.

Figure F.11: Local Getis-Ord Gi* values for modal adverbs in preverbal position.
Figure F.12: Local Getis-Ord Gi* values for modal adverbs in postverbal position.


