



University of Pennsylvania Working Papers in Linguistics

Volume 22

Issue 2 *Selected Papers from New Ways of Analyzing
Variation (NWAY 44)*

Article 13

12-1-2016

The Role of Geography in Syntactic Variation: A Corpus-based Analysis of Adverb Position across Varieties of English Worldwide

Edwin Ko

The Role of Geography in Syntactic Variation: A Corpus-based Analysis of Adverb Position across Varieties of English Worldwide

Abstract

Recent investigations have shown an effect of geography on syntactic variation among dialects in a handful of languages. Developments in corpus-based dialectology have introduced sophisticated analysis techniques for studying dialect differences. 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. Following the findings of Grieve 2012, which identified the dialects of the Northeast against those in the Southeast and South Central states by examining the positions of adverbs from a corpus of written Standard American English, this study employs a similar approach to spoken and written corpora of World Englishes, drawing from the International Corpus of English (ICE). Applying spatial autocorrelation techniques to identify similarities and differences due to geographical space shows significant overall spatial clustering in the spoken data, but not in the written data. English varieties in India, Hong Kong, Singapore and the Philippines were identified as 'hot spots' for placing certain adverbs before the verb, whereas varieties in Canada and Ireland acted as hot spots for positioning adverbs after the verb.

The Role of Geography in Syntactic Variation: A Corpus-based Analysis of Adverb Position across Varieties of English Worldwide

Edwin Ko*

1 Introduction

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., for North American English, Kurath 1972, Labov et al. 2006; for British English, Szmrecsanyi 2013; and for Dutch, Spruit 2008, among many others). As dialect corpora are becoming increasingly available, developments in corpus-based dialectology have introduced sophisticated 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, that embraces these kinds of computational and statistical tools to study variation due to geographical space.

Still not much is known about the role of geography in 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 any role to play in similarities or differences among World Englishes. Hence if differences among World Englishes are found, what are the likely mechanisms underlying this variation?

The present study considers adverb position by extracting placement frequencies from spoken and written corpora of ten varieties of English worldwide. By employing spatial analysis techniques, global Moran's I and local Getis-Ord G_i^* , this study aims to discover whether spatial patterning exists for adverb placement across a variety of World Englishes. It is found that the spoken data, but not the written data, exhibits significant spatial clustering for several variables, demonstrating that regional patterning manifests in non-contiguous varieties. In addition, the (South-) East Asian and East African varieties tend to pattern in a geographically similar way, as do the English varieties in Canada, Jamaica and Ireland. These findings raise various questions regarding the source of variation in adverb position among World Englishes. This study also complements existing work on the relationship between geography and syntactic variation through the use of naturalistic corpora and statistical techniques.

2 Background

2.1 Adverb Placement

It is generally agreed that certain adverbs may change in meaning when placed in a different syntactic position (e.g., Cinque 1999, Ernst 2002, Jackendoff 1972, Morzycki 2015, among many others). For example, consider the following constructions containing the adverb *cleverly* in which its meaning is sensitive to its placement within the sentence:

- (1) a. Alice *cleverly* has answered the questions.
- b. Alice has *cleverly* answered the questions.
- c. Alice has answered the questions *cleverly*.

In (1a), there is a single meaning of *cleverly*, in that Alice was clever for answering the question (i.e., subject-oriented). In (1c), there is also a single meaning of *cleverly*, but the meaning is differ-

*Many thanks to Amir Zeldes, Jack Grieve, and the participants at NWA V 44 for helpful comments on this project. All errors are my own.

ent from (1a). Here, the adverb has a manner reading and can be paraphrased as *Alice has answered the questions in a clever way*. Finally, the sentence in (1b) is ambiguous between the readings found in (1a) and (1c).

Several scholars have also identified adverbs which may be placed in a different position without altering their meaning. Among them is Jackendoff (1972:42) who notes that adverbs such as *quickly, slowly, reluctantly, sadly, quietly, indolently, frequently, immediately, often* and *soon* do not have a “discernible change in meaning.” Ernst (2002) further notes that pure manner adverbs are unable to obtain a clausal reading, and thus can only have a manner reading, which is illustrated by the following two sentences:

- (2) a. Jordan *loudly* spoke.
 b. Jordan spoke *loudly*.

According to Ernst pure manner adverbs include adverbs such as *loudly, brightly, limply, tightly, woodenly* and other adverbs that “often require an event as involving a physical stimulus, such as volume of sound” (2002:87).

2.2 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 they depended on the judgments of the dialectologist, and were therefore difficult to replicate. The work by Séguy (1971) was seminal in introducing the methodology known as dialectometry which, according to Szmrecsanyi (2011), is “the branch of linguistics 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 (2011:45). Thus, the notion of ‘corpus-based’ is simply the integration of, and dependence on, corpus data, and the marriage between corpus linguistic methodology and statistical methods.

In his work, Grieve (2012) employs two spatial autocorrelation techniques, global Moran's I and local Getis-Ord G_i^* , in order to identify (a) whether adverb position exhibits any regional patterning, and (b) whether there are specific locations that act as ‘hot spots’ for patterning within written Standard American English. In one study, he analyzes ‘common’ adverbs (see the appendix of Grieve 2012 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 (or non-sentence-initial), 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. In addition, these findings were based on written documents which Schneider (2002) claimed to conceal much of the author's speech pattern, thereby limiting the amount of regional or sociolinguistic variation in these registers.

3 Methodology

Data for this study were from the International Corpus of English (ICE),¹ which consists of several different varieties of English spoken worldwide, but only ten of these varieties were analyzed: Northern Ireland and the Republic of Ireland (ICE-Ireland), Kenya and Tanzania (ICE-East Africa), 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 component and 400,000 words in the written component). The corpora also follow a common design. It is im-

¹See <http://ice-corpora.net/ice>.

portant to note that while the ICE-Ireland and ICE-East Africa corpora contain a million words, these corpora have been divided so that each of the four varieties end up containing 500,000 words in total for both registers.

Three adverb positions (sentence-initial, pre-verbal and post-verbal), based on Jackendoff’s (1972) seminal work on adverb placement, and four adverb classes (evaluative, e.g., *honestly*; frequency/temporal, e.g., *often*; manner, e.g., *loudly*; and modal, e.g., *probably*) were chosen for this study.² Each corpus was tokenized and tagged for their parts-of-speech using the NLTK Toolkit (Bird et al. 2009), and all tokens tagged as adverbs were extracted from across the spoken and written corpora. The most common adverbs were then manually selected for each adverb class based on the examples provided for each adverb class by Cinque 1999, Geuder 2000 and Ernst 2002, and also based on my own judgments. Then, if a token matched one of the adverbs within the four adverb 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. 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 post-verbal 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, or (d) the next token is either a preposition, conjunction, or punctuation. If an adverb did not meet the criteria for either sentence-initial or post-verbal position, then it is in pre-verbal position. One hundred sentences in each corpus were checked manually for correct placement of adverbs, and of 1,000 sentences across all corpora, the script achieved more than 95% accuracy.

The distribution of adverbs in spoken and written texts is displayed in Figure 1 and 2. Surprisingly, the distributions in both texts appear to be very similar and so there does not seem to be any noticeable difference in how adverbs are placed between the two registers. Sentence-initial position is generally disfavored across all adverb classes, especially manner adverbs which almost categorically do not appear sentence-initially. Manner adverbs in pre- and post-verbal position also have similar distributions. Furthermore, pre-verbal position is a common placement among adverbs with the exception of evaluative adverbs, which are placed in all three positions at around the same rate (in comparison to other adverb classes), with a preference for post-verbal position.

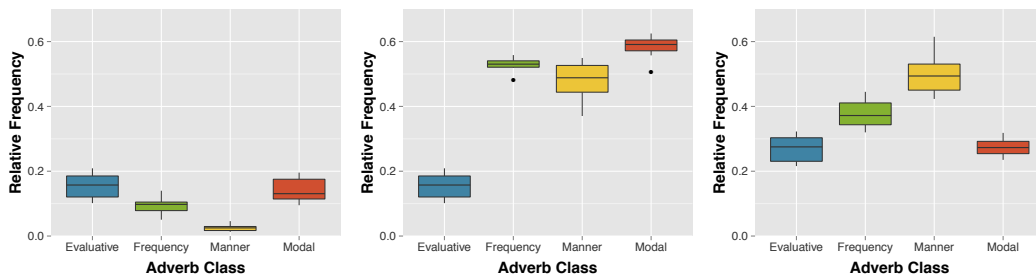


Figure 1: Distribution of adverbs sentence-initially (left), pre-verbally (center), and post-verbally (right) within the spoken texts.

²Initially, the class of manner adverbs consisted only of pure manner adverbs. However, pure manner adverbs appeared very infrequently within the texts. In the spoken texts of Canadian and Tanzanian English, these adverbs only occurred seven times across the three positions. The class of manner adverbs was therefore expanded to also include aspect-manner (e.g., *quickly*) and resultative adverbs (e.g., *beautifully*) based on the subtle nuances in their meaning (see Ernst 2002 and Geuder 2000 for further discussion on these adverbs).

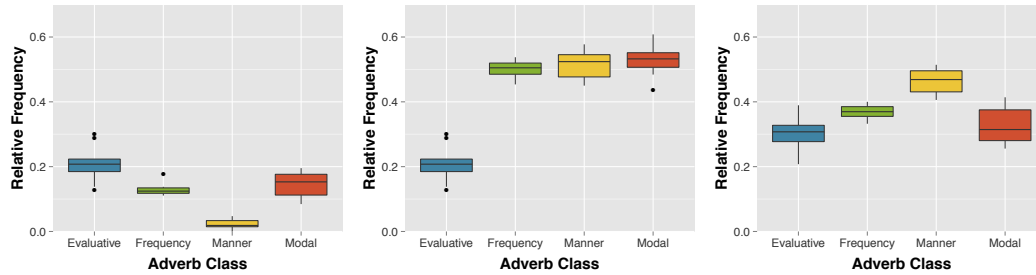


Figure 2: Distribution of adverbs sentence-initially (left), pre-verbally (center), and post-verbally (right) within the written texts.

The spatial analysis techniques employed in this study were global Moran's I and local Getis-Ord G_i^* using the `moran.test` and `localG` functions of the `spdep` package for R (Bivand et al. 2015). Global Moran's I 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. Local Getis-Ord G_i^* 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 tend to 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 G_i^* 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.

The calculated Moran's I value falls within $[-1, 1]$, in which a significant positive value represents an overall clustering. To determine the significance of Moran's I, the p-value must be less than or equal to 0.016 for $\alpha = 5\%$ (or greater than or equal to a corresponding z-score of 2.14 for a one-tailed p-value because the interest here is whether there is any spatial clustering). This p-value is adjusted using Bonferroni correction ($0.05 / 3 = 0.016$) since more than one position within each adverb class is being analyzed and compared. It is important to mention that a binary spatial weighting function was used with a cutoff distance of 6,000km, although a cutoff of 3,000km and 8,550km were tested as well. The spatial weighting function allows pairs of locations that are closer together to receive more weight than locations that are farther apart. For local Getis-Ord G_i^* , the p-value was also 0.016, but since we are interested in whether a location exhibits clustering of either low or high values, a two-tailed p-value was used, which corresponds to a z-score of ± 2.41 . The local Getis-Ord G_i^* returns a z-score for each location, whereas the global Moran's I returns only the p-value.

4 Results

The analysis of global spatial autocorrelation identified five significant variables: post-verbal frequency/temporal adverbs, pre- and post-verbal manner adverbs, and modal adverbs in sentence-initial and pre-verbal positions. Frequency/temporal adverbs in pre-verbal position also exhibited near-significant spatial clustering. Interestingly, none of the variables in the written texts were found to display significant geographical patterning.

The map in Figure 3 displays local Getis-Ord G_i^* values for frequency/temporal adverbs in pre-verbal position across the ten locations in the spoken data. Global Moran's I reveals that this variable exhibits near-significant positive spatial autocorrelation ($I = 0.36$, $p = 0.018$). In other words, frequency/temporal adverbs in pre-verbal position demonstrated near-significant spatial clustering overall. Turning to the local Getis-Ord G_i^* scores, none of the locations displayed significant spatial clustering of either low or high values.

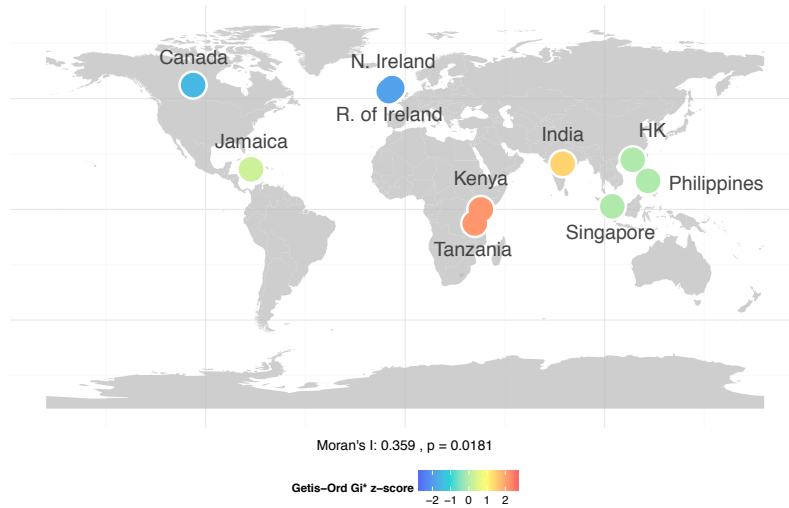


Figure 3: Local Getis-Ord G_i^* values for frequency/temporal adverbs in pre-verbal position in the spoken texts.

In Figure 4, the local Getis-Ord G_i^* values for post-verbal frequency/temporal adverbs in the spoken data are mapped out. Moran's I was 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 G_i^* values are all negative. In contrast, the G_i^* values for varieties in Canada, Jamaica, Northern Ireland, and the Republic of Ireland are all positive. Moreover, the G_i^* 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 post-verbal position, and speakers of English in locations nearby Ireland tend to also place adverbs after the verb.

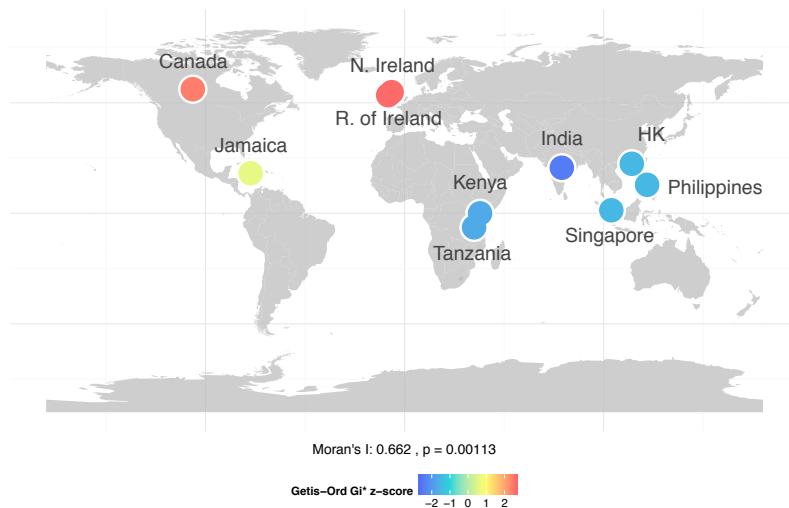


Figure 4: Local Getis-Ord G_i^* values for frequency/temporal adverbs in post-verbal position in the spoken texts.

Figure 5 maps the local Getis-Ord G_i^* values for manner adverbs in pre-verbal position in the spoken data. According to global Moran's I for this variable, a significant positive spatial autocor-

relation was found ($I = 0.71$, $p < 0.01$). Turning to the G_i^* scores, we see 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. The G_i^* values further show that Canada, Northern Ireland and the Republic of Ireland exhibited significant spatial clustering of low values, whereas India displayed significant spatial clustering of high values.

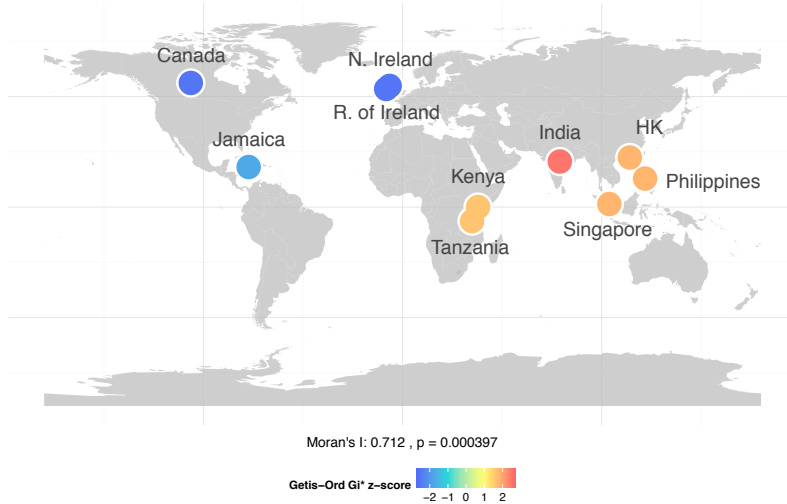


Figure 5: Local Getis-Ord G_i^* values for manner adverbs in pre-verbal position in the spoken texts.

Figure 6 shows the local Getis-Ord G_i^* values for manner adverbs in post-verbal position in the spoken data. Global Moran's I reveals significant positive spatial autocorrelation ($I = 0.63$, $p < 0.01$). 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 G_i^* scores are opposite to what we observed in Figure 5, which displays results for manner adverbs but in pre-verbal position. Northern Ireland and the Republic of Ireland were identified as significant hot spots (i.e., exhibiting local spatial autocorrelation) of high values for this variable.

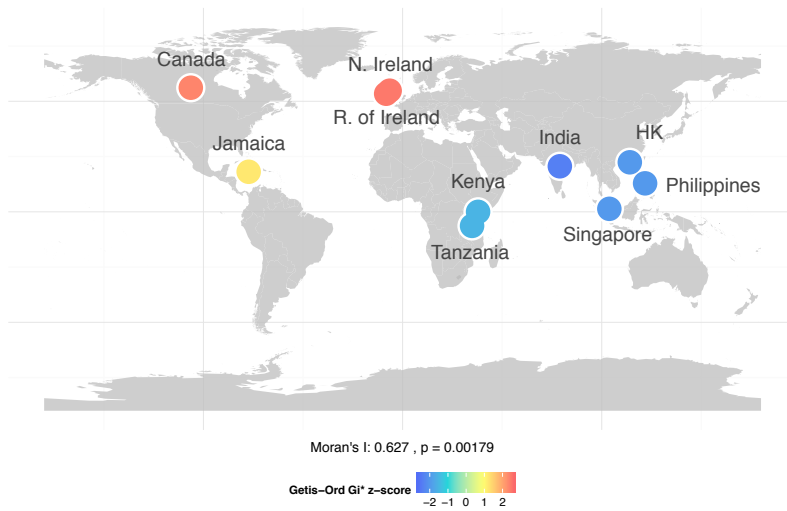


Figure 6: Local Getis-Ord G_i^* values for manner adverbs in post-verbal position in the spoken texts.

In Figure 7, the map displays the local Getis-Ord G_i^* scores for modal adverbs in sentence-initial position across the ten locations in the spoken data. Moran's I was found to exhibit significant positive spatial autocorrelation ($I = 0.53, p < 0.01$). None of the locations exhibited significant spatial clustering of low or high values.

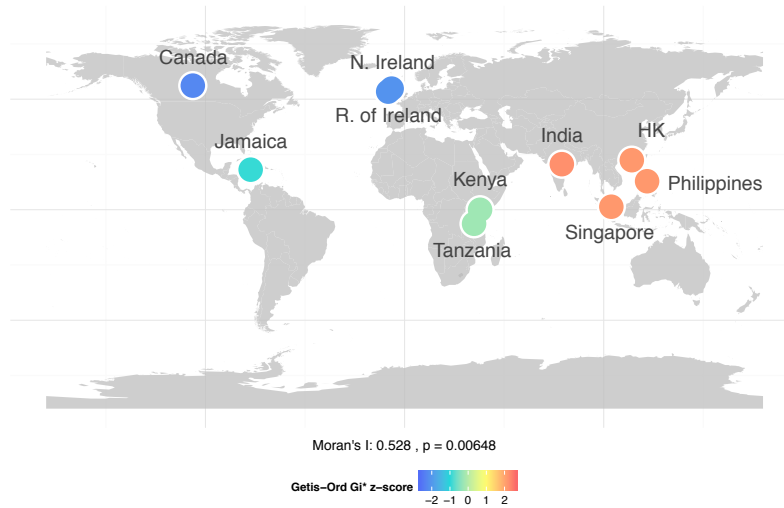


Figure 7: Local Getis-Ord G_i^* values for modal adverbs in sentence-initial position in the spoken texts.

Finally, the map in Figure 8 displays the local Getis-Ord G_i^* values for modal adverbs in pre-verbal position across the ten locations in the spoken data. Moran's I was found to be significant, displaying positive spatial autocorrelation ($I = 0.54, p < 0.01$). 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 G_i^* 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. Nevertheless, Singapore, Hong Kong and the Philippines were shown to exhibit significant spatial clustering of low values.

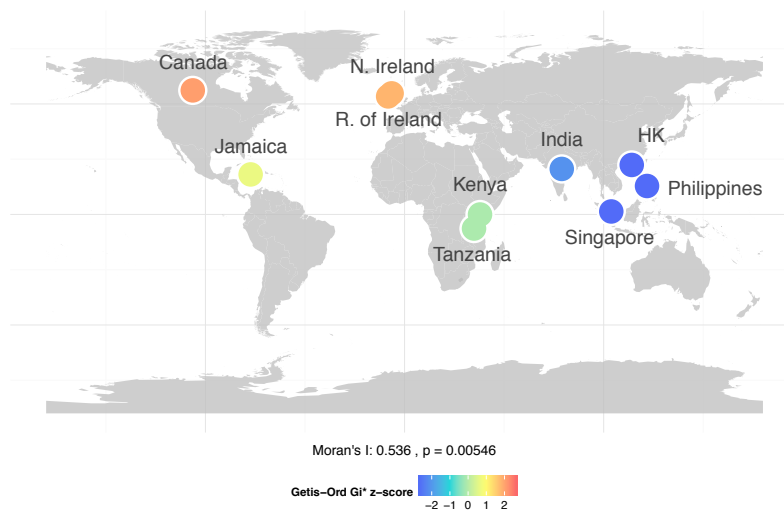


Figure 8: Local Getis-Ord G_i^* values for modal adverbs in pre-verbal position in the spoken texts.

5 Discussion

The analysis of global spatial autocorrelation identified six variables: pre- and post-verbal frequency/temporal adverbs, manner adverbs in pre- and post-verbal position, and modal adverbs in sentence-initial and post-verbal 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 indicates that adverb placement exhibits a weaker (geo)linguistic 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.

In comparing the local Getis-Ord G_i^* scores across the significant variables, two recurring trends emerge. First, Canadian, Jamaican, and the Northern/Southern Irish Englishes tend to pattern in a similar way. For example, in the spoken data, Ireland was identified as a hot spot of low values for manner adverbs in pre-verbal position and a hot spot of high values in post-verbal position. Moreover, Canada and Jamaica share similar z-scores with Ireland, and this pattern was consistent among other significant variables. The second trend is that varieties in East Africa and in (South-)East Asia also tend to pattern in a similar way.

These results suggest that Canadian, Jamaican and the Northern/Southern Irish Englishes form a single cluster in terms of their similarities in a placement preferences for adverbs, and the East African and (South-)East Asian Englishes form another cluster. These categorizations are similar to Kachru's (1992) Three Circle model of English: the Inner Circle (e.g., English-as-a-Native-Language varieties), the Outer Circle (e.g., English-as-a-Second-Language varieties), and the Expanding Circle (e.g., English-as-a-Foreign-Language varieties). Since none of the varieties analyzed in this study are considered EFL varieties, only the former two are of interest here. While there have been discussions that criticize the fuzziness of the boundaries of this tripartite model (e.g., Galloway and Rose 2015, Jenkins 2015), the typological distinctions that the model puts forth (i.e., ENL and ESL, or native and non-native), nevertheless, appear to be strong indicators of how speakers of these ten English varieties tend to pattern in terms of their adverb placement.

Critics have also noted that Kachru's Three Circles of English do not take into account bi- and multilingualism which are integrally interwoven in many of these societies (e.g., Bruthiaux 2003). In Canada, English and French are recognized as official languages, and in addition to Jamaican English, Jamaican Creole forms an important part of Jamaican identity (Schneider 2007). Therefore, it is worth considering not only the proficiencies of the speakers, but also the possibility for contact-induced changes and interferences that may have arisen due to factors driven by intense contact, and bi- and multilingualism. In fact, research in second-language acquisition has suggested that adverb placement is difficult to acquire even for advanced learners (White 1991). In these studies, they find that L1 language transfer effects play a prominent role in the acquisition of adverb placement. As such, second-language acquisition strategies may be important mechanisms by which these placement tendencies emerge especially in locales where English is learned as a second language.

If adverb placement in the contact language(s) parallels the distribution of adverb placement in English, then contact may have a significant influence on the speakers' use of English. For example, in Irish, adverbs and adverbials appear after the verb (McCloskey 2011). It is possible then that Irish may have influenced the preference for post-verbal position among English speakers in Ireland which was identified as a hot spot of high values for post-verbal adverbs. In India, which exhibited spatial clustering of high values for English manner adverbs in pre-verbal position, multilingualism is highly pervasive, affecting many aspects of life in the country. Most of the official languages of India (i.e., languages that belong to the 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 has the highest population of speakers 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 therefore be prone to syntactic transfer effects from their first language (Thomason 2001).

In either case, 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. Thus, we would expect local languages spoken in India and languages spoken in Singapore to be structurally more similar, and languages spoken in India and languages spoken in Canada to be more dissimilar. 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. (2007).

Despite the significant findings of this study, there are several important limitations I would like to point out. First, several manner adverbs that were analyzed exhibited slight differences in meaning. There are several reasons why this should not significantly affect the results of this study. The first reason is that in many of the locales studied here (e.g., Hong Kong), English is generally spoken as a second language; the lower proficiency may mean that the subtle nuances in meaning are not properly distinguished in L2. Secondly, in locations where English is spoken natively, many speakers are also proficient in another language (e.g., French-English bilinguals in Canada) which introduces the possibility for cross-linguistic influence. The third reason is that there is still no general consensus as to which adverbs remain semantically equivalent. While Jackendoff (1972) lists *quickly* as an adverb that does not appear to change in meaning, Ernst (2002) classifies it as an aspect-manner adverb and notes the subtle differences in its meaning. Among pure manner adverbs, Ernst includes *tightly*, an adverb that could be argued to belong to the class of resultative adverbs (e.g., *beautifully*) that also appear to have subtle differences in their meaning (Geuder 2000). Lastly, there are several studies where meaning differences have not been controlled for, yet their findings have made important contributions (e.g., White 1991).

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 by adding 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 because these texts do not follow the same common design employed by the ICE corpus. The corpus of Global Web-based English (Davies and Fuchs 2015), a 1.9-billion-word corpus featuring twenty World English varieties, is a potential next step to address the issues that follow from a limited sample size, and to determine whether these results generalize to other World Englishes.

Finally, the corpus design may also pose problems. Apart from the corpora of Kenyan, Tanzanian, Northern Irish and Southern Irish English, which consist of 500,000 words each, the other corpora contain one million words. All of these corpora are further divided into spoken and written components. This means that the East African and Northern/Southern Irish varieties have half the number of words in the spoken and written texts in relation to the other corpora. While the decrease in corpus size lowers the reliability, the underlying distribution of adverbs is assumed to be the same. Furthermore, these corpora were compiled during different periods and while they follow a 'common' design, it is up to the compilers how they interpret these guidelines. Thus, temporal and genre variation may play a role in the outcome of this study.

6 Conclusion

This study has shown that adverb placement exhibits strong spatial patterning among ten (non-) contiguous varieties of World Englishes. In the spoken texts, 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. No significant patterning was found

within the written texts, supporting Schneider's (2002) claim that written registers conceal much of an author's speech patterns. 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. These findings contribute to the growing body of research in geolinguistics, and open up an abundance of research avenues into variation and change across World Englishes from a variety of perspectives.

References

- Bird, Steven, Ewan Klein and Edward Loper. 2009. *Natural Language Processing with Python*. O'Reilly Media.
- Bivand, Roger, Micah Altman, Luc Anselin, Renato Assunção, Olaf Berke, Andrew Bernat, Guillaume Blanchet, Eric Blankmeyer and Marília Carvalho. 2015. Package 'spdep.' R package, version 0.5-92.
- Bruthiaux, Paul. 2003. Squaring the circles: Issues in modeling English worldwide. *International Journal of Applied Linguistics* 13:159–178.
- Cinque, Guglielmo. 1999. *Adverbs and Functional Heads: A Cross-linguistic Perspective*. New York: Oxford University Press.
- Davies, Mark and Robert Fuchs. 2015. Expanding horizons in the study of World Englishes with the 1.9 billion word Global Web-based English Corpus (GloWbE). *English World-Wide* 36:1–28.
- Ernst, Thomas. 2002. *The Syntax of Adjuncts*. Cambridge, UK: Cambridge University Press.
- Galloway, Nicola and Heath Rose. 2015. *Introducing Global Englishes*. New York, NY: Routledge.
- Geuder, Wilhelm. 2000. Oriented Adverbs: Issues in the Lexical Semantics of Event Adverbs. Doctoral dissertation, Universität Tübingen.
- Grieve, Jack. 2012. A statistical analysis of regional variation in adverb position in a corpus of written Standard American English. *Corpus Linguistics and Linguistic Theory* 8:39–72.
- Grieve, Jack. 2016. *Regional Variation in Written American English*. Cambridge, UK: Cambridge University Press.
- Holman, Eric, Christian Schulze, Dietrich Stauffer and Søren Wichmann. 2007. On the relation between structural diversity and geographical distance among languages: Observations and computer simulations. *Linguistic Typology* 11:393–422.
- Jackendoff, Ray. 1972. *Semantic Interpretation in Generative Grammar*. Cambridge, MA: MIT Press.
- Jenkins, Jennifer. 2015. *Global Englishes: A Resource Book for Students*. Abingdon, UK: Routledge.
- Kachru, Braj. 1992. World Englishes: Approaches, issues and resources. *Language Teaching* 25:1–14.
- Kurath, Hans. 1972. *Studies in Area Linguistics*. Bloomington, IN: Indiana University Press.
- Labov, William, Sharon Ash and Charles Boberg. 2006. *Atlas of North American English: Phonetics, Phonology and Sound Change*. Berlin: Mouton de Gruyter.
- McCloskey, James. 2011. The shape of Irish clauses. *Formal Approaches to Celtic Linguistics*, 143–178.
- Morzycki, Marcin. 2015. *Modification*. Cambridge, UK: Cambridge University Press.
- Schneider, Edgar. 2002. Investigating variation and change in written documents. *The Handbook of Language Variation and Change*, 67–96.
- Schneider, Edgar. 2007. *Postcolonial English: Varieties Around the World*. Cambridge, UK: Cambridge University Press.
- Séguy, Jean. 1971. La relation entre la distance spatiale et la distance lexicale. *Revue de Linguistique Romane* 35:335–357.
- Sharma, Aryendra. 1975. *A Basic Grammar of Modern Hindi (Spoken and Written)*. New Delhi: Central Hindi Directorate.
- Spruit, Marco. 2008. Quantitative Perspectives on Syntactic Variation in Dutch Dialects. Doctoral dissertation, University of Amsterdam.
- Szmrecsanyi, Benedikt. 2011. Corpus-based dialectometry: A methodological sketch. *Corpora* 6:45–76.
- Szmrecsanyi, Benedikt. 2013. *Grammatical Variation in British English Dialects: A Study in Corpus-Based Dialectometry*. Cambridge, UK: Cambridge University Press.
- Thomason, Sarah. 2001. *Language Contact: An Introduction*. Washington, DC: Georgetown University Press.
- White, Lydia. 1991. Adverb placement in second language acquisition: Some effects of positive and negative evidence in the classroom. *Second Language Research* 7:133–161.