Tobler’s first law of geography: "everything is related to everything else, but near things are more related than distant things" (Tobler 1970).

The big question: Does adverb placement exhibit geographical patterning in varieties of English around the world, and if so, what is their likely source?

1 Introduction

- Many researchers have observed an influence of geographical distance on syntactic variation in dialects from a handful of languages (Spruit et al. 2009; Grieve 2012).

- Developments in corpus-based dialectology have introduced sophisticated analysis techniques for studying dialect differences due to geographical space.
  - In particular, spatial autocorrelation techniques are used to identify spatial clustering of certain linguistic features, which have been employed in studies in geography for several decades.

- Grieve (2012) analyzes regional patterning of adverb placement in written Standard American English via use of spatial autocorrelation techniques:
  - Global Moran’s I (Moran 1948): examines whether a variable exhibits clustering or not.
  - Local Getis-Ord Gi* (Ord & Getis 1995): identifies which locations are hot spots of high/low value clustering.

- By using spatial autocorrelation techniques, do we also find geographical patterning of adverb placement among varieties of English around the world?

2 Data

- The data comes from the International Corpus of English (ICE) which includes the following countries/cities: Ireland (Northern Ireland & Republic of Ireland), East Africa (Kenya & Tanzania), Jamaica, Hong Kong, India, Philippines, Singapore, and Canada.

- Each corpus contains a million words (600,000 in the written texts and 400,000 in the spoken texts), and follows a similar corpus design.

*Many thanks to Amir Zeldes, Jack Grieve, Paul Portner, Cathy O’Connor, and Mark Sicoli for helpful comments on various stages of this project, and to Keith Ord for comment on the statistics and results. Thanks also goes to the anonymous reviewers who provided very helpful remarks. All errors are my own.
Corpora of English varieties spoken in Northern Ireland, Republic of Ireland, Kenya and Tanzania contain half a million words each.

- Longitudes and latitudes coordinates were obtained from NASA’s latitude/longitude finder (2015), and the population statistics of each location (in the year the corpus was being constructed) were obtained from The World Bank (2015).

3 Method

3.1 Adverbs

- Adverbs may change in meaning when placed in a different syntactic position. For example:
  - Sentence-initial position: *Cleverly, Jordan wrote the book.*
    - The adverb "cleverly" contains subject-oriented reading which can be paraphrased as *it was clever of Jordan to have wrote the book.*
  - Pre-verbal position: *Jordan cleverly wrote the book.*
    - The adverb "cleverly" may either contain a manner or subject-oriented reading. The manner reading describes the clever way in which Jordan wrote the book.
  - Post-verbal position: *Jordan wrote the book cleverly.*
    - The adverb "cleverly" contains a manner reading.

- Following the classification of adverbs proposed by Ernst (2002), this study analyzes four adverb classes that do not change in meaning.
  - Evaluative: (un)fortunately, inevitably, admittedly, evidently, surely, etc.
  - Frequency: regularly, recently, usually, normally, gradually, etc.
  - (Pure) Manner: quickly, loudly, smoothly, beautifully, gently, etc.
  - Modal: probably, perhaps, possibly, (un)likely, definitely, etc.

3.2 Corpus Analysis

- Each corpus was cleaned, tokenized, and tagged for parts-of-speech using Python scripts and the Natural Language Toolkit (Bird et al. 2009).

- Adverbs were hand picked to ensure that they belonged to one of the four adverb classes and did not change in meaning.

- Three adverb positions were analyzed: Sentence-Initial, Pre-Verbal, and Post-Verbal.

- Probabilities were extracted from each corpus for each adverb class in each adverb position.

- Written and spoken texts were analyzed separately.

1In order to ensure the scripts were predicting the adverb positions correctly, a hundred sentences were manually checked from each corpus.
3.3 Spatial Analysis

- Euclidean syntactic distances of each adverb class were calculated for each pair of locations and modeled using a Mantel test (1967)\(^2\) to test for correlation between:
  
  (a) geographical distance, which was calculated based on the longitude and latitude coordinates.
  
  (b) Trudgill’s (1974) notion of linguistic gravity, a measure incorporating geographical distances with population statistics.\(^3\,\,\,4\)

- Global Moran’s I and local Getis-Ord Gi*, were used to identify spatial clustering globally, and which locations exhibit clustering of high/low values.

- The global and local spatial autocorrelation were calculated with a cutoff distance of 8550km, which resulted in an average of 4.6 links between the pairs of locations. A reciprocal weighting function was also used.\(^5\)

4 Results: Spoken Data

Written data was not found to display any geographical patterning.\(^6\)

4.1 Adverb Placement Distributions

Figure 1: Distribution of adverbs across all spoken corpora in sentence-initial position (left), pre-verbal position (center), post-verbal position (right).

- Adverbs generally prefer to be placed pre-verbally.

- Manner adverbs have greatest preference for post-verbal position in comparison to other adverb classes.

- Manner adverbs appear split between pre-verbal and post-verbal positions.

- Written data exhibit similar adverb distributions, but with less spread (i.e. lower range).

---

\(^2\)Mantel test is a non-parametric statistic that tests for correlation between two non-independent distance matrices.

\(^3\)The equation for Trudgill’s Linguistic Gravity Index (TGLI) is calculated using the formula \(I_{ij} = \log \frac{P_i P_j}{d_{ij}^2}\), where \(I_{ij}\) represents the logged mutual influence between the two population centers, \(P_i\) is the population of location \(i\), etc., and \(d_{ij}\) is the distance between locations \(i\) and \(j\) (cf. Nerbonne and Heeringa 2007).

\(^4\)The idea is that interaction between two populations will correlate positively as distance between them decreases, but correlate negatively as distance increases.

\(^5\)It should be noted that, given the small number of locations, these results would be sensitive to the weights linking different locations. Other cutoff distances such as 3000km and 6000km would provide better insights on the nature of these weights.

\(^6\)All results will refer to analyses on the spoken data. Results regarding written will be explicitly stated.
4.2 Mantel Test

Figure 2: Euclidean syntactic distances of frequency adverbs correlated with geographical distance (left), of manner adverbs correlated with Trudgill’s Linguistic Gravity Index TGLI (center), and of modal adverbs correlated with TGLI (right).

- Geographical distance explains 24.0% of the variance among frequency adverb placement at $p = 0.03$.
- TLGI explains 11.9% of the variance among manner adverb placement at $p = 0.03$.
- TLGI explains 30.5% of the variance among modal adverb placement at $p = 0.04$.
- Euclidean syntactic distances of modal adverbs are generally low – the differences in modal adverb placement among the ten locations are slight in comparison to those in manner or frequency adverb placement (c.f. Fig. 1).

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

Figure 3: Local spatial autocorrelation of manner adverbs in pre-verbal position. Moran’s I = 0.295, $p < 0.01$. Significant clustering of high values around Singapore and clustering of low values around Canada and Jamaica ($p < 0.016$, adjusted using Bonferroni correction: $\alpha / n = 0.05 / 5 = 0.016$, where $n$ refers to the number of variables being analyzed).
Three variables in the spoken data were found to display significant geographical clustering overall: pre-verbal placement of manner adverbs (e.g. *loudly sang*), post-verbal placement of manner adverbs (e.g. *sang loudly*), and post-verbal placement of frequency adverbs (e.g. *left recently*).\(^7\)

For manner adverbs, the ‘hot spot’ of high values for pre-verbal position was analyzed around Singapore, while a ‘hot spot’ of low values was analyzed around Canada. In other words, English speakers in Singapore tend to place adverbs pre-verbally over other positions, more so than Canada, where speakers tend to avoid placing adverbs pre-verbally.

The opposite was observed for post-verbal positions for manner and frequency adverbs. English speakers in Canada prefer placing adverbs post-verbally, while English speakers in Singapore disfavor placing adverbs post-verbally.

The dichotomy in manner adverb placement suggests that sentence-initial position has little to say about the variation in placement of manner adverbs.

\(^7\)A valid concern is the small number of locations which will influence the power of the test. Keith Ord, via personal correspondence, said that “I am impressed and perhaps a little surprised that your analysis produced such strong results; that said, your results appear to be fine.” Nevertheless, one should be cautious when interpreting results using a small sample size.
5 Discussion

Are the similarities and dissimilarities in adverb placement in these varieties of English due to other varieties of English that are geographically proximate, or are they due to other factors such as language contact?

5.1 Shift-Induced Change

- Kroch (2001): "One actuating force for syntactic change whose existence cannot be doubted is language contact."

- In most of these nations, virtually all people speak a language other than English as a first language.

- If a language is learned as a second language, then it would be more prone to shift-induced inference such as syntactic transfer effects due to imperfect learning (Thomason 2001).

- It has been noted that adverb placement in English is difficult to acquire, even for advanced learners (White 1991, Chan 2004).

- Syntax tends to be resilient to change (especially since the history of English in some of these locations is still relatively young), but more malleable when it comes to multilingualism.

- In Chinese, which is spoken by the majority of the population in Singapore, adverbs are usually placed before the verb (Li and Thompson 1981, Po-Ching and Rimington 1997).

- In French, which is widely spoken in Canada, post-verbal placement in non-finite contexts is favored by native speakers via preference/grammaticality judgment tasks (Ayoun 1999).

- If these preferences arose due to contact with the local languages, why is there a gradual shift in preferences of adverb placement in English across the globe?

5.2 Typological similarity and geographical distance

- 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 aren’t 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, Singapore and India are more geographically proximate than, say, Singapore and Canada. Because of this, we would expect local languages spoken in Singapore and languages spoken in India to be more structurally similar. The effects may be diametrically opposite for local languages spoken in Singapore and languages spoken in Canada. As a consequence of contact, English varieties worldwide will then mirror the structures of the more dominant languages spoken in their respective countries resulting in a similar geographical patterning reported in the study by Holman et. al (2007).
If these results are borne out, they might actually reflect Tobler’s (1970) first law of geography: “everything is related to everything else, but near things are more related than distant things”.

6 Future Research

• Extension of these spatial analysis techniques to adverb placement in these contact languages to determine if we obtain the same geographical patterning.

• Preference/grammaticality judgment tasks on adverb placement in English varieties as well as the contact languages to get a better understanding of whether syntactic transfer effects are actually happening.

• Replication of this study using the Global Web-Based English Corpus, which includes English varieties spoken in 20 countries/cities – this study only looks at 10 different locations which is admittedly low for these kinds of spatial analysis techniques.

• Aggregation of other linguistic features such as genitive alternations (e.g. the boy’s book vs. the book of the boy) to see if these features also exhibit geographical patterning, and if so, do contact languages provide an explanation?

• What are the processing and contextual factors that influence adverb placement (c.f. Szmrecsanyi et al. to appear) and how do they interact with changes due to contact?

7 References


