

Dialect typology: Recent advances

Melanie Röthlisberger (University of Zurich) and Benedikt Szmrecsanyi (KU Leuven)

Corresponding author: Melanie Röthlisberger
Englisches Seminar
Plattenstrasse 47
8032 Zürich
Switzerland
melanie.roethlisberger@es.uzh.ch

Dialect typology: Recent advances

Melanie Röthlisberger (University of Zurich) and Benedikt Szmrecsanyi (KU Leuven)

1. Introduction

Research in DIALECT TYPOLOGY (also known as SOCIOLINGUISTIC TYPOLOGY) lies at the interface of LINGUISTIC TYPOLOGY, which is concerned with categorizing human languages based on their structural differences and similarities, and DIALECTOLOGY, which is the study of regionally or socially defined and typically vernacular forms of language. The intersection between typology and dialectology has received considerable attention in recent years (see Kortmann 2004; Szmrecsanyi and Wälchli 2014). Common to research in dialect typology is interested in the “extent to which differences of linguistic structure, whether within or between languages, can be ascribed to or explained in terms of features of the society in which the dialects in question are spoken” (Trudgill 1996:3; Trudgill 2004; Trudgill 2009a; Trudgill 2011). The aim of this chapter is to survey recent work in this spirit. It is important to note that the extant literature on dialect typology has mostly concentrated on varieties of English, which thanks to their geographical and typological diversity offer an exciting testing ground for theories about dialect typology. The current article – which is a substantially extended version of Szmrecsanyi and Röthlisberger (forthcoming) – inherits this bias towards varieties of English, but care will be taken to highlight insights from the dialect-typological literature that can be generalized to other dialect and variety landscapes.

This chapter is structured as follows. In Section 2, we set the stage by discussing a number of geographical factors that are assumed in the literature to have a bearing on the structural make-up of different languages and dialects such as world region, geography and contact with speakers of other languages or dialects. Section 3 sketches the extent to which dialects of languages exhibit common features (e.g. “vernacular universals” à la Chambers 2004). In Section 4, we synthesize work on parameters of structural diversity of languages, namely analyticity versus syntheticity (i.e. the degree to which grammatical information is conveyed by suffixation), and complexity versus simplicity, all within the context of geographic space. Section 5 offers some concluding remarks.

2. Geographical factors

2.1. World region and geography

The regional distribution of linguistic features takes center stage in both areal typology and in classical dialectology (see Murelli and Kortmann 2011), and it is thus not surprising that researchers have explored areal effects extensively. As such, research has focused on whether varieties or dialects spoken in a geographically bounded area, that is, bounded for instance by mountain ranges, rivers, or the sea, share a set of features that are absent from varieties spoken elsewhere (for instance, varieties of British English versus varieties of American English). For instance, in the realm of World Englishes, studies have shown that West African varieties tend to have five-vowel systems, that British varieties show a preference for an extensive system of diphthongs (Schneider 2004:1127), and that some specific phonological features are particularly diagnostic of regional accents, for instance yod-dropping or the TRAP vowel (Schneider 2004:1129). Regarding grammar, we know from surveys (e.g. Kortmann and Szmrecsanyi 2004) that American varieties of English have special forms or phrases for the second person plural pronoun (e.g. *you all*); that British varieties tend to have existential / presentational *there's, there is, there was* with plural subjects; that Asian varieties of English exhibit irregular (from a standard English perspective) use of articles; that

Caribbean varieties attest, for example, multiple negation; and that African varieties tend to use a wider range of the progressive.

Studies that take an interest in geography as a decisive factor to classify dialects have often aimed to correlate dissimilarity between varieties with the varieties' geographical distance. In general terms, the areal null hypothesis asserts that geographic proximity between dialects or varieties should predict linguistic similarity between these dialects and varieties (Nerbonne and Kleiweg 2007:154 refer to this as the "Fundamental Dialectology Principle"). In this line of research, linguistic similarity is gauged by assessing the distribution of linguistic features in a predefined area. This distribution, obtained from different localities, is then clustered into distinct groups that are taken to designate dialect areas. For instance, Szmrecsanyi (2012a) compares 30 native varieties of English based on 235 morphosyntactic features sampled in the *World Atlas of Varieties of English* (WAVE) (see Kortmann and Lunkenheimer 2012 and <http://ewave-atlas.org/>). To visualize the similarities and differences between those varieties, the paper presents a NeighborNet diagram, which is reproduced in Figure 1 below. NeighborNet diagrams were originally developed in biometry and bioinformatics to map phylogenies and reticulate effects such as genetic recombination. They have become quite popular in historical linguistics, in cross-linguistic typology (e.g. Dunn et al. 2008) and in dialectology (e.g. McMahon et al. 2007). Without insisting on a strictly phylogenetic interpretation, Figure 1 visually depicts aggregate similarities and distances between varieties of English. The diagram can be read like a family tree that is not rooted. Branch lengths are proportional to linguistic distance: Distance in the plot broadly corresponds to morphosyntactic dissimilarity, that is, the closer two varieties are in the plot and the shorter the line that connects them, the more similar the varieties are with respect to their morphosyntactic make-up based on the 235 linguistic features in WAVE. The most important split in Figure 1 is between so-called 'high-contact' (Trudgill 2009b) native varieties at the bottom of the diagram and the other native varieties in the sample (more on the distinction between high-contact and low-contact varieties in Section 2.2. below). In other words, the degree of contact a variety is/was exposed to exerts the greatest influence to distinguish varieties along morphosyntactic lines. The bottom of the diagram further includes a second areal cluster consisting of Rural African American English, Urban African American English, and South Eastern American English, that is, those varieties are structurally very similar to each other. The majority of the remaining varieties are clustered at the top of the diagram and include some areal sub-groupings, such as Welsh/Southwest/North English (a British cluster) and Colloquial American/Ozarks /Appalachian English (an American cluster).

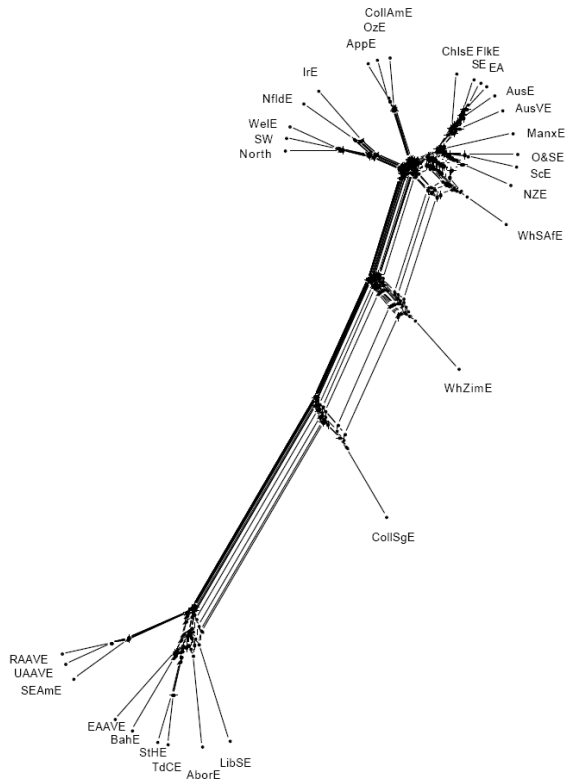


Figure 1. Visualizing aggregate similarities between 30 native varieties of English with a NeighborNet diagram. Distances (branch lengths) are proportional to cophenetic linguistic distances (adapted from Szmrecsanyi 2012a:Figure 4). A list of abbreviations is provided in the appendix.

The clustering of varieties based on their linguistic similarity and dissimilarity is an excellent method if one is interested in determining the geographic boundaries of larger dialect areas. Including hitherto under-researched syntactic variation in their dialectometric analysis of Swiss-German dialects, Scherrer and Stoeckle (2016) cluster information obtained on 350 variables from 377 locations in German-speaking Switzerland and find homogeneously formed dialect regions (Scherrer and Stoeckle 2016:106). Boundaries between dialect regions thereby partly coincide with political borders and topography. Calculating the arithmetic mean, standard deviation and skewness of the data for each linguistic level by location, they further show that the syntactic level provides less clear-cut dialect boundaries compared to the other linguistic levels, that Zurich and Bern constitute the center of two internally coherent dialect areas, and that the Valais, south of the Alps' mountain ridges, develops linguistically independently of the northern dialects and constitutes a more conservative dialect area (Scherrer and Stoeckle 2016:115–116).

Finally, quantifying the correlation between linguistic similarity and geographic proximity can give an indication of the importance of geography compared to other factors. To that end, pairwise linguistic distances (as calculated by, for example, the number of feature classifications in the *World Atlas of Varieties of English* with regard to which two varieties differ) are correlated with pairwise geographic distances (as the crow flies, in km, calculated using a standard trigonometry formula). For the set of varieties of English shown in Figure 1, the correlation coefficient between morphosyntactic and geographic distances amounts to $r = .226$ ($p < .001$). This means that while there is a significant areal relationship between varieties, geographic distance only explains 5.1% ($R^2 = 0.051$) of the morphosyntactic variability in the dataset (see Gooskens 2004 for similar low correlations in Norwegian dialect data). This is a modest share compared to dialectometric measurements in traditional English dialects. Shackleton (2007), for instance, observes that geographic distance

accounts for about 49% of phonetic variation in traditional English dialects in England. While these differences in explanative value could be ascribed to differences in dialect types (World Englishes vs. traditional dialects), this discrepancy also indicates that the various levels of language do not correlate to the same extent with geographic distances (see Scherrer and Stoeckle 2016:94). Spruit et al. (2009), for instance, observe that phonological and syntactic features correlate much better with geographic distance than lexical features in their analysis of Dutch dialect data, possibly due to internal inconsistencies in their lexical dataset or due to structural constraints which do not apply to lexical data. This asymmetry runs counter to Montemagni (2008) who found morpho-lexical distances to correlate better with geographic distances than phonetic distances in Tuscan dialects. Similarly, Scherrer and Stoeckle (2016) observe a higher correlation between morphological dissimilarity and geographic distance than between syntactic dissimilarity and geographic distance in their analysis of Swiss German dialects. Focusing on the same dialect area, Jeszenszky et al. (2017) use residuals obtained from regression models as a proxy for correlation between syntactic variation and geographic distances (see Jeszenszky et al. 2017:95). Their results indicate that the correlation between syntactic distance and regional differences is higher than reported for phonological variation (e.g. Nerbonne 2010) and additionally confirm the results in Spruit (2008) on Dutch dialect data in that the correlation between geographic and syntactic distances is better described in terms of a linear relationship than a logarithmic one. What is more, the findings in Jeszenszky et al. (2017) indicate that geographic distance cannot explain linguistic variance as well as travel times ($R^2 = 0.458$ vs. $R^2 =$ between 0.775 and 0.783) which seems to be a better proxy of language contact. These results further confirm Szmrecsanyi's (2012b) observation that the impact of geographic distance might be overrated as an explanans. In sum, these, partly contradictory, findings indicate that the type of linguistic features, the kind of data sampled (see Jeszenszky et al. 2017:105) and the dialect area examined play a crucial role in determining the importance of geography for the classification of dialects.

2.2. Exposure to language and dialect contact

Typological work with an interest in cross-linguistic comparisons has often focused on the impact that language contact has on the systematic distribution of structural features across the world's languages (see Aikhenvald and Dixon 2001; Siemund and Kintana 2008). Research in that spirit aims to disentangle the principles and mechanisms of contact-induced change, and pays particular attention to the constraints (linguistic as well as social) that exert influence on the contact-based structural similarities between languages (Siemund 2008:3). Constraints that have taken center stage in pertinent scholarship include the architecture and prestige of the languages involved, the degree of bilingualism, the length of contact, the number of speakers, and numerous other parameters (see Siemund 2008:4). Research in areal typology (see previous section) has further shown that these structural similarities can also arise between genetically unrelated languages in what has been called "sprachbund" or "linguistic areas" (see Matras 2009:236, 266).

The typological literature on the principles and mechanisms involved in language contact has stimulated a good deal of dialectological inquiry. Pioneering this line of research was Peter Trudgill, who argues that language and dialect contact was and is the driving force for the emergence and diversification of new varieties, at least with regard to English (see Trudgill et al. 2000; Trudgill 2006; Trudgill 2008). Trudgill (2009c:320) distinguishes between two types of varieties based on their degree of language contact: "Low contact" varieties are long-established mother tongue dialects with a low degree of contact with other dialects or languages. "High contact" varieties, on the other hand, include all varieties where speakers have been and still are in extensive contact with speakers

of other languages. It is particularly this later type that is of interest for studies on language and dialect contact. “High contact” varieties include:

- Non-native indigenized L2 varieties (e.g. Indian English)
 - transplanted L1 varieties or (post-)colonial standards (e. g. New Zealand English)
 - language-shift varieties (e. g. Irish English)
 - Standard L1 varieties (e.g. British English)
 - Creoles (e.g. Hawai’i Creole)
- (see also Kortmann and Szmrecsanyi 2011:15f.)

While it is agreed that new varieties frequently emerge out of contact with other dialects and languages, predicting the structural outcome of such contact is a challenging task (Siemund 2008:3). There is, however, overall consensus that two possible outcomes of contact-induced change ought to be distinguished, namely complexification and simplification (see Section 4.2). The binary distinction of these two processes (and their dichotomy) has recently been challenged. Taking a diachronic perspective on the development of Tristan da Cunha English – an isolated variety of English spoken in the south Atlantic (Schreier (2016). He concludes that the degree of typological similarity between the languages/dialects in contact impacts the structural outcome of dialect or language contact more than the extent to which the two varieties are in contact (Schreier 2016:145). That is, simplification arguably only occurs in high-contact situations if the two linguistic systems are maximally different from each other. The typological dissimilarity between linguistic systems in high-contact scenarios is a crucial factor to account for different outcomes of dialect (same language) and language (different languages) contact. According to Schreier, language contact settings lead to simplification while dialect contact settings do not.

This contrast between language and dialect contact-induced change is valuable if one were to account for distinctions in language acquisition processes: In dialect contact scenarios, the majority of language learners are children who select various features from the heterogeneous input of the feature pool (Trudgill 2010). Arguably then, childhood language acquisition results in an increase of linguistic variants and complexification, as in the case of New Zealand English, Canadian English, or American English. On the other hand, language contact scenarios entail adult learners acquiring a second language, which inevitably leads to simplification as a result of the limited language acquisition abilities of adult learners (e.g. as in the case of India, Singapore, or Hong Kong English) (Trudgill 2010). Based on the premise that we find simplification in those varieties where a language is spoken by adult learners and signs of complexification in those varieties where a language has been acquired by (bilingual) children, we can predict more simplification in indigenized L2 varieties of English compared to colonial L1 varieties. Whether this hypothesis matches the facts will be discussed in Section 4.2.

The influence of language and dialect contact on the degree of similarity or dissimilarity between varieties has furthermore been linked to the topography of the geographic area in which the languages under investigation are spoken. Topography can be defined in terms of physical reality, such as vegetation, stream flow, length of growing season, availability of water, climate, altitude and barriers (see Nichols 2013:38). These aspects shape linguistic diversity and the diffusion (or non-diffusion) of linguistic features. On the basis of a case study of languages of the Nakh-Daghestanian family, spoken in the mountain regions of the Caucasus, Nichols (2013) illustrates that linguistic innovations spread uphill from the (peripheral) lowlands to the more isolated and conservative (but situated in the geographic center) highlands through seasonal workers and trading routes. This situation then led to “asymmetrical vertical bilingualism” with altitude correlating with linguistic structural properties (Nichols 2013:54). A similar impact of topography on linguistic diffusion is observed by Jeszenszky et al. (2017) who report that syntactic distance correlates statistically

significantly less well with geographic distance (calculated with Euclidean distance) ($r = 0.65$) than with travel times in 2000 ($r = 0.744$), 1950 ($r = 0.743$), and 1850 ($r = 0.737$) in German-speaking Switzerland (Jeszenszky et al. 2017:98, 105). On the local level, however, this impact of topography, defined as the difference between as-the-crow-flies distance and travel times, bears different roles depending on region: In the mountainous area around the Bernese Oberland including the Alpine ridges to the Valais, travel times is a much better predictor of linguistic distance (from $r=0.815$ in 1850 to $r=0.674$ in 2000) than geographic distance ($r=0.445$) (Jeszenszky et al. 2017:101) and the difference in effect is greater than on the global level. Conversely, areas that exhibit more homogeneous topography also show less difference between the correlations of geographic distance and travel times with linguistic distance presumably due to increased dialect contact (Jeszenszky et al. 2017:102).

2.3 Variety type

Finally, varieties of a language can further be distinguished based on their sociohistorical background. In the literature on World Englishes, the most basic typology customary in the literature (see also Crystal 2004; Kortmann and Lunkenheimer 2013) distinguishes the following variety types¹:

- **Native L1 varieties** (such as e.g. Canadian English). This type roughly corresponds to the Inner Circle in Kachru (1992).
- **Indigenized L2 varieties** (such as e.g. educated Jamaican English). This type roughly corresponds to the Outer Circle in Kachru (1992).
- An inclusive typology will also recognize **pidgin and creole languages based on a particular lexifier** (such as Tok Pisin, which is English-lexified), as a third type.

This typology is essentially defined in terms of whether or not we are dealing with a contact language and with regard to how and when the variety is acquired (first language acquisition versus second language acquisition). Against this backdrop, dialect typologists have been primarily concerned with establishing those linguistic features that are particularly diagnostic of specific types. Such features have been collectively referred to as “varioversals” (Szmrecsanyi & Kortmann 2009a), that is, “features recurrent in language varieties with a similar socio-history, historical depth, and mode of acquisition” (33) (see also Section 3 below).

In order to identify varioversals empirically, researchers often make use of survey databases, similar to the World Atlas of Language Structures (WALS) (Dryer & Haspelmath 2013) or the Atlas of Pidgin and Creole Language Structures Online (APiCS; <http://apics-online.info>; see Michaelis et al. 2013). Such survey databases have become an indispensable tool in the field of cross-linguistic typology. One such popular survey database is the morphosyntax survey (<http://www.varieties.mouton-content.com/>) that accompanies the *Handbook of Varieties of English* (Kortmann et al. 2004). This survey of non-standard English morphosyntax includes a catalogue of 76 non-standard features covering 46 vernacular varieties of English around the world (see Kortmann & Szmrecsanyi 2004 for discussion). Information on the attestation of these features was obtained from various linguistic experts on the varieties. Each expert was asked to rate the features in the relevant variety according to the following categories:

¹ We acknowledge that an argument could be made to include Learner varieties in this typology. However, in keeping with much of the dialect typology literature we take the liberty to not consider Learner varieties.

- A pervasive (possibly obligatory) or at least very frequent
- B exists but a (possibly receding) feature used only rarely, at least not frequently
- C does not exist or is not documented

The survey reveals that distinctive L1 varioversals (i.e. morphosyntactic features that are particularly characteristic of L1 varieties of English) include existential / presentational *there's, there is, there was* with plural subjects (e.g. *There's two men waiting in the hall*); *me* instead of *I* in coordinate subjects (e.g. *Me and my brother*); and adverbs having the same form as adjectives (e.g. *Come quick!*). The top L2 varioversals include lack of inversion in main clause yes/no questions (e.g. *You get the point?*); irregular use of articles (e.g. *Take them to market, I had nice garden, about a three fields, I had the toothache*); and levelling of the difference between the present perfect and the simple past (e.g. *Were you ever in London?, Some of us have been to New York years ago*). And finally, varioversals characteristic of English-lexified pidgin and creole languages comprise lack of inversion / lack of auxiliaries in lack of inversion / lack of auxiliaries in *wh*-questions (e.g. *What you doing?*); lack of inversion in main clause yes/no questions (e.g. *You get the point?*); and special forms or phrases for the second person plural pronoun (e.g. *youse, y'all, aay', yufela, you ... together, all of you, you ones/'uns, you guys, you people*).

While the distinction between native and non-native varieties (and thus a variety's sociohistorical background) is still predominantly used for the classification of varieties of English, variety type has also been of main concern in creolistic studies for the classification of creoles versus non-creoles. Perez (forthcoming), for instance, compares Portuguese- and Spanish-lexified creoles with postcolonial dialects of Spanish and Portuguese (including alleged semi-creoles such as Afro-Brazilian Portuguese and Caribbean Spanish) and the two lexifiers – European Spanish and Portuguese – on the basis of the attestation of 60 morphosyntactic, lexical, and phonological features sampled from the *Atlas of Pidgin and Creole Structures (APiCS)*, Michaelis et al. 2013), *WALS* and the descriptive literature. While the study primarily aims to offer a classification of Afro-Yungueño Spanish whose status as creole remains contested (see Lipski 2008; Sessarego 2013; Perez 2015), the study's findings also suggest that same-lexifier creoles are structurally more similar to one another and at the same time maximally different from their lexifier languages. Figure 2 illustrates the structural similarity and dissimilarity between varieties with a NeighborNet diagram shown in Perez (forthcoming). The length of the branches between varieties is proportional to structural similarity: the shorter the connecting lines, the more similar the two varieties and vice versa. The diagram plots creoles to the left, including semi-creoles, and the standard varieties and Spanish / Portuguese dialects to the right. Features that are only shared by creoles include lack of gender distinction, verbal suppletion according to tense or tense and aspect, lack of nominal plural markers, overt subject marking (note that Spanish and Portuguese are pro-drop languages), frequent lack of definite article, lack of or invariant copula, and polar questions in situ with or without particle (see Perez forthcoming).

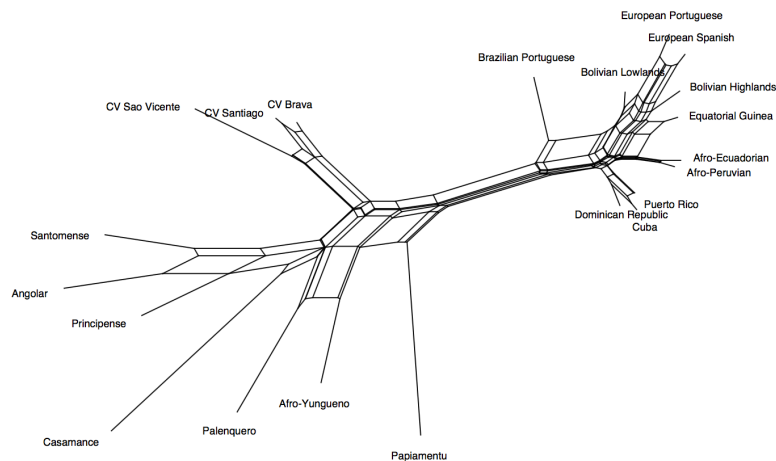


Figure 2. NeighborNet diagram of Spanish- and Portuguese-lexified creoles, dialects of Spanish and Portuguese and standard varieties.

3. Dialect universals, implications, and related notions

The quest for generalizations, also known as *linguistic universals*, has been dominating research in cross-linguistic typology for much of the second half of the twentieth century. This quest has been shown to be reconcilable with the presumed contrasting interest in areal diffusion (see Bickel 2017:40). Szmrecsanyi and Kortmann (2009a:33) present the following typology of -versals:

- (i) GENUINE UNIVERSALS (e.g. *all languages have vowels*);
- (ii) TYPOVERSALS, i.e. features that are common to languages of a specific typological type (e.g. *SOV languages tend to have postpositions*);
- (iii) PHYLOVERSALS, i.e. features that are shared by a family of genetically related languages (e.g. *languages belonging to the Indo-European language family distinguish between masculine and feminine gender*);
- (iv) AREOVERSALS, i.e. features common to languages which are in geographical proximity to each other (e.g. *languages belonging to the Balkan Sprachbund have finite complement clauses*);
- (v) VERNACULAR UNIVERSALS, i.e. features that are common to spoken vernaculars (e.g. *spoken vernaculars tend to have double negation*);
- (vi) features that tend to recur in vernacular varieties of a specific language: ANGLOVERSALS, FRANCOVERSALS, etc. (e.g. *in English vernaculars, adverbs tend to have the same morphological form as adjectives*);
- (vii) VARIOVERSALS, i.e. features recurrent in language varieties with a similar socio-history, historical depth, and mode of acquisition (e.g. *L2 varieties of English tend to use resumptive pronouns in relative clauses*).

Note that these -versals are occasionally all presented as universals although it is questionable to what extent (ii) to (vii) are genuine universal features, as their names already imply. Rather, types of -versal can be observed in languages or varieties of languages that have some sociohistorical aspects in common.

Classic typologists tend to be concerned with (i) to (iv), while (v) to (vii) comes under the remit of dialect typology. Dialect typologists have been particularly interested in (v): VERNACULAR UNIVERSALS. Chambers (e.g. Chambers 2004) defines VERNACULAR UNIVERSALS as “a small number of phonological

and grammatical processes [that] recur in vernaculars wherever they are spoken [...] not only in working class and rural vernaculars, but also in [...] pidgins, creoles and interlanguage varieties” (2004:128). For English, Chambers has proposed the following features as candidates for vernacular universalhood (Chambers 2004:129):

- (ng) or alveolar substitution in final unstressed *-ing*, as in *walkin’, talkin’* and *runnin’*.
- (CC) or morpheme-final consonant cluster simplification, as in *pos’ office*, *han’ful*.
- final obstruent devoicing, as in *hundret* (for *hundred*), *cubbert* (for *cupboard*).
- conjugation regularization, or leveling of irregular verb forms, as in *Yesterday John seen the eclipse* and *Mary heard the good news*.
- default singulars, or subject-verb nonconcord, as in *They was the last ones*.
- multiple negation, or negative concord, as in *He didn’t see nothing*.
- copula absence, or copula deletion, as in *She smart* or *We going as soon as possible*.

As Chambers himself notes, the examples might be from English but since they are “primitive features, not learned” (and thus part of the language faculty), they are not considered to be restricted to English only (Chambers 2004:129).

The quest for universals has recently experienced a turn towards an interest in the underlying stochastic constraints that shape linguistic variation. To the extent that two or more varieties share the same probabilistic grammar, they are said to be similar. Probabilistic grammars can be defined as the underlying structure that drives speakers’ linguistic choice-making. Building on the comparative sociolinguistic approach advanced by Poplack and Tagliamonte (2001) and Tagliamonte (2002), similarities or dissimilarities between varieties’ probabilistic grammars are then quantified based on the statistical significance of constraints, the constraints’ relative importance and their effect size. For such a comparison, separate regression models or random forest are computed per variety and differences in the output of these models (e.g. coefficient estimates from the regression models, ranking of predictors from the random forest) are submitted to dialectometric analyses applying Multidimensional Scaling (MDS). MDS (Kruskal & Wish 1978) is a well-known dimension reduction technique that translates distances between objects (in our case, language varieties) in high-dimensional space into a lower-dimensional representation. Conducting a study in this vein on particle placement (*He looked up the word* versus *He looked the word up*) in varieties of English, Grafmiller and Szmrecsanyi (in press) show that there is a high degree of stability among L1 varieties of English while L2 varieties are more dissimilar to themselves as well as to the L1 varieties. Their results suggest that “the patterns we see in our data are at least partly attributable to biases in L2 acquisition”. Adapting the same methodology to an analysis of the dative alternation (e.g. *Mary gave John the apple* versus *Mary gave the apple to John*) in World Englishes, Röthlisberger (2018) finds a cluster of American English-influenced varieties (Canadian English and Philippine English) that contrasts on the probabilistic level with British English-influenced varieties (British English, Hong Kong English, Indian English, Irish English, Jamaican English, New Zealand English and Singapore English). The results of that multidimensional scaling analysis are shown in Figure 4. The two dimensions (x-axis and y-axis) account for 89.1% of the variance in the data. Proximity between varieties is taken as a sign of similarity in probabilistic grammars while distance between varieties is taken as a sign of dissimilarity in probabilistic grammars.

Linguistic universals, both on the surface level as well as in the probabilistic domain, might thus provide the grounds on which the typological relatedness between different variety types can be assessed.

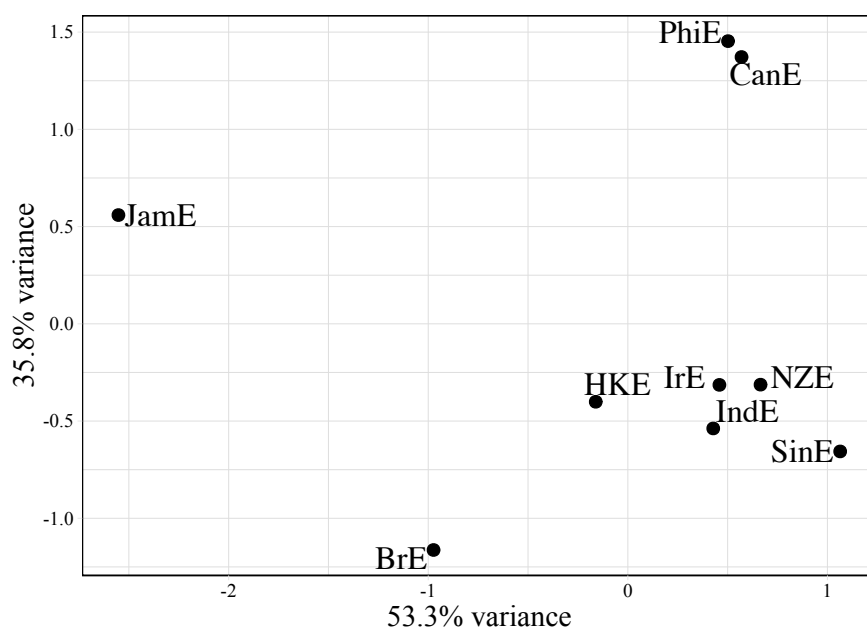


Figure 4. Multidimensional scaling map of nine varieties of English. Distances between varieties correspond to differences in the coefficient estimates obtained from by-variety logistic regression models including the five most important predictors (see Röthlisberger 2018:170). (See the appendix for a list of abbreviations.)

Researchers in pursuit of universals are also often interested in co-occurrence patterns of linguistic features in order to gain a better understanding of the evolution of such features. Such co-occurrence patterns can be biconditional implications (for instance, ‘if in a language the genitive follows the noun, then the complement follows the adposition, and vice versa’; Greenberg 1963), or one-way implications (for instance, ‘if a language has a marked singular, it has also a marked plural, but not necessarily vice versa’; Greenberg 1966). With regard to World Englishes, for instance, Szmrecsanyi and Kortmann (2009b) find that 94% of the varieties covered in the *Handbook of Varieties of English* either have both *ain’t* as the negated form of *have* (e.g. *I ain’t had a look at them yet*) as well as *ain’t* as the negated form of *be* (e.g. *They’re all in there, ain’t they?*), or they have neither. Needless to say, this biconditional implication is consonant with the dialectological literature (Anderwald 2003:149–150). Investigating co-occurrence patterns from a more quantitative perspective, Szmrecsanyi (2017) applies multiple correspondence analysis (MCA) (Lê, Josse and Husson 2008; Levshina 2015:375–376) to co-occurrence patterns in the morphosyntax survey of the *Handbook of Varieties of English*. With that methodological tool under his belt, Szmrecsanyi (2017) investigates the extent to which linguistic features are associated with each other and characteristic of certain varieties: A particular variety will appear in the same part of the plot as the values of the features by which the variety is characterized. The World Englishes MCA plot is shown in Figure 5.

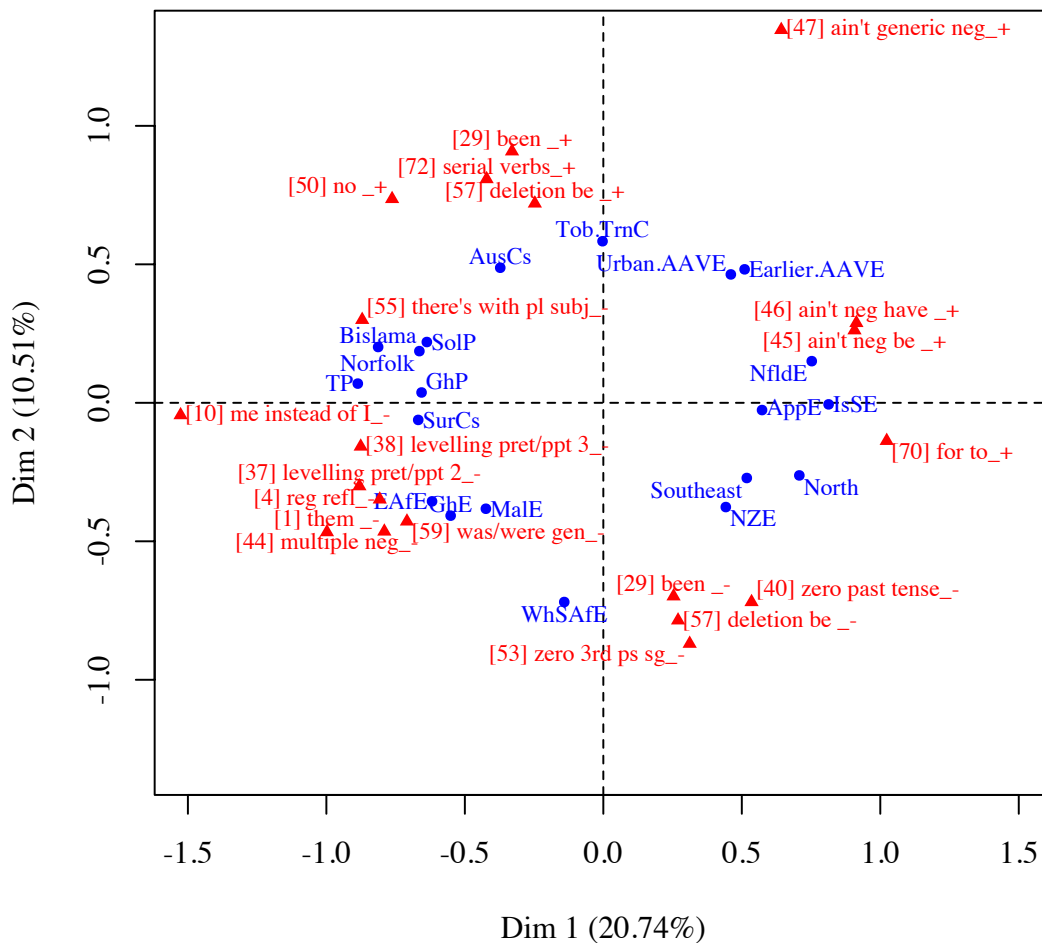


Figure 5. Multiple Correspondence Analysis (MCA) map, based on the morphosyntax survey coming with the *Handbook of Varieties of English*. Proximity between features indicates co-occurrence patterns. Display is limited to the 20 features and varieties that have the highest contribution on the dimensions. '+' suffixed to a feature's label indicates presence of the feature, '-' indicates absence. (source: Szmrecsanyi 2017:Figure 2)

Figure 5 can be summarized as follows: The upper left-hand quadrant contains features such as serial verbs and *no* as preverbal negator, which are demonstrably characteristic of English-based pidgin and creole languages; and indeed, the varieties that the MCA plot identifies as particularly attracted to these features (e.g. Australian Creoles) are all pidgins and creoles. The upper right-hand quadrant plots three features as particularly distinctive: *ain't* as generic negator before a main verb, *ain't* as the negated form of *have*, and *ain't* as the negated form of *be*. This co-occurrence pattern ties in with what was said above. It should be furthermore noted that this quadrant includes only North American varieties, indicating that – as is well-known – *ain't* is particularly characteristic of North American Englishes. In the lower right-hand quadrant the plot locates some British varieties as well as New Zealand English, a variety that is known to be relatively close to British English, at least in terms of grammar. Features that are characteristic of these varieties include e.g. unsplit *for to* in infinitival purpose clauses, while e.g. deletion of *be* is typically absent. This distributional pattern is typical of British varieties of English (Kortmann and Szmrecsanyi 2004:1162–1165). In the lower left-

hand quadrant, we find mainly indigenized L2 varieties such as Ghanaian English. As suggested in Figure 5, distinctive features associated with these varieties include the absence of features such as *them* instead of demonstrative *those*, and multiple negation. In sum then, the following dialect typology is suggested by the MCA: the most important dimension of variation (Dim 1) pits native varieties (right) against pidgins/creoles and L2-varieties (left). The vertical dimension (Dim 2) appears to be capturing a language-externally defined contrast between orientation toward British English (bottom) versus orientation toward North American English (top). (Note also that this is consonant with the findings in Röthlisberger 2018). This pattern nicely illustrates how dialect-typological work needs to consider both language-internal and language-external factors to uncover useful generalizations.

4. Parameters of structural diversity

Similarities between languages and dialects are often discussed in terms of the variety or language's sociohistorical background, the degree of contact (low vs. high), or shared linguistic (morphosyntactic) features (e.g. varioversals) as elucidated above. The present section will introduce two additional sets of parameters that capture the structural diversity of varieties: analyticity vs. syntheticity (section 4.1.) and complexity vs. simplicity (section 4.2).

4.1. Analyticity versus syntheticity

The distinction between analytic and synthetic languages originates in the works of August Wilhelm von Schlegel (1818) and Sapir (1921) who proposed a number of parameters along which languages should be categorized into broader types. Greenberg (1960), based on Sapir's typology, later defined five indices to characterize languages (Greenberg 1960:185). Inspired by Greenberg's work, Szmrecsanyi (2009) and Szmrecsanyi and Kortmann (2009c) apply Greenberg's cross-linguistic methodology to the study of a geographically widespread range of varieties of English. They investigate the degree of grammatical analyticity and syntheticity by contrasting the frequency of free vs. bound grammatical markers per word. *Formal grammatical analyticity* captures all coding strategies where grammatical information is encoded with free grammatical markers defined as closed-class function words without any lexical meaning. In contrast, *formal grammatical syntheticity* includes all coding strategies where grammatical information is encoded with bound grammatical markers (see Szmrecsanyi 2009:2).

The degree of a variety's analyticity or syntheticity due to language contact has been directly linked to processes of simplification and complexification (see also Section 2.3). Generally speaking, simplification is assumed to be the result of widespread adult Second Language Acquisition (SLA) (see Trudgill 2010) and can manifest itself in two possible ways: regularization of irregularities and an increase in lexical and morphological transparency (Trudgill 2010:307; but see Hengeveld and Leufkens 2018). Lexical and morphological transparency entails analytic structures where "the relation between form and meaning is as transparent as possible" and "every single meaning is expressed in a separate form" (Kusters 2003:21). The direct relation between form and meaning facilitates language acquisition by adult speakers (Trudgill 2010:312). The most extreme cases where language contact results in simplification (and hence increased analyticity) are pidgins and creoles (see also Leufkens 2013; McWhorter 2001). In contrast, low contact situations arguably lead to complexification and increased syntheticity. Evaluating these hypotheses empirically with data taken from two English-lexified creoles (Tok Pisin and Hawai'i Creole), from a number of rural dialects of British English, non-native indigenized L2 varieties, transplanted L1 varieties, and language-shift varieties, Siegel et al. (2014) observe that creoles encode grammatical information significantly less often synthetically than analytically compared to other varieties of English. At the same time, their results indicate that creoles are not more analytic in absolute terms than L2 or L1 varieties of English. Locating varieties in a two-dimensional syntheticity-analyticity space, Siegel et al. (2014)

show that both Tok Pisin and Hawai'i Creole use synthetic markers less often than other varieties of English while exhibiting a similar degree of analyticity at the same time. This is illustrated in Figure 6: the x-axis plots the degree of syntheticity, with the number of synthetic features increasing as one moves from left to right in the figure; the y-axis plots the degree of analyticity, with the number of analytic features increasing as one moves from the bottom to the top of the figure. Figure 6 thus also indicates that L1 varieties and traditional British dialects use both more synthetic and analytic markers compared to some L2 varieties where zero marking is comparatively frequent (see Kortmann & Szmrecsanyi 2011:275).

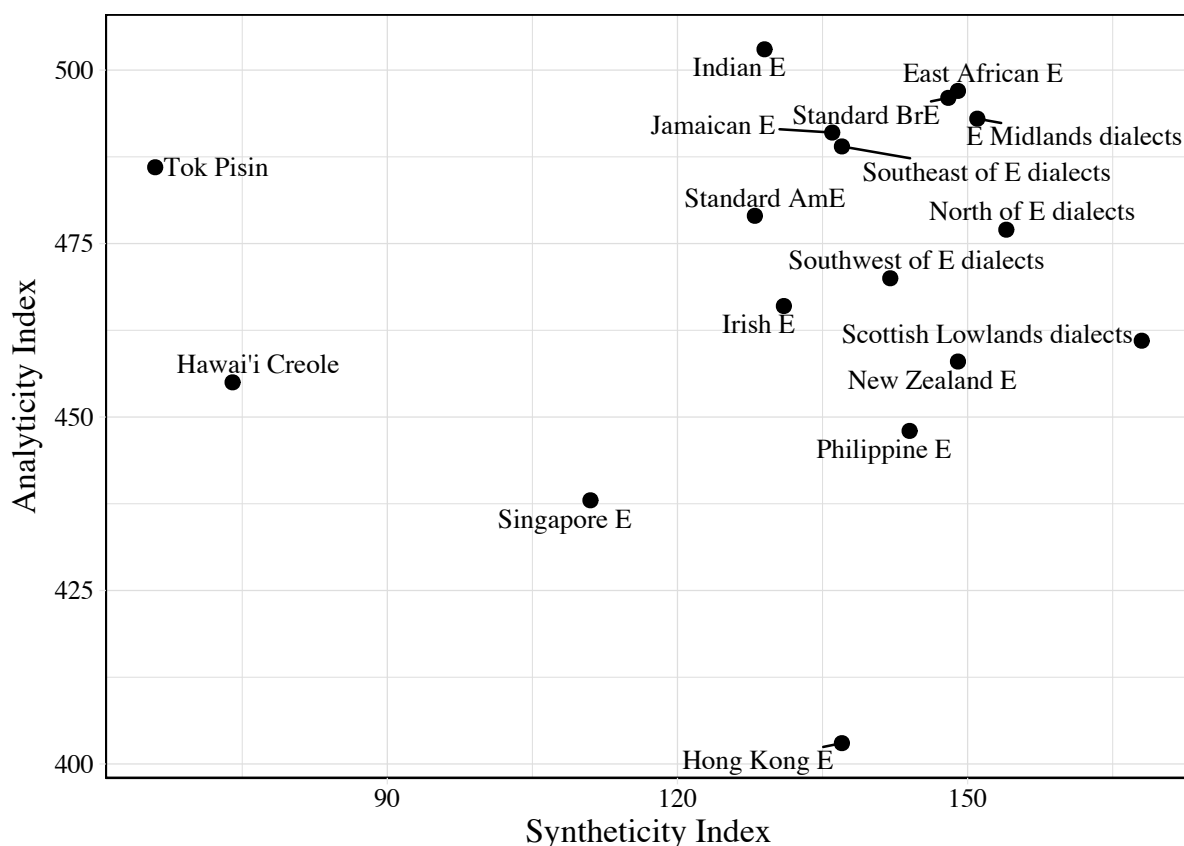


Figure 6. Tok Pisin and Hawai'i Creole vis-à-vis varieties of English: total number of analytic types against total number of synthetic types (focus on inventory sizes). (adapted from Siegel et al. 2014: Figure 2)

In a similar spirit, studies that compare structural innovations in learner and L2 varieties often observe a preference for morphological transparency in both variety types compared to native varieties. For instance, results in Callies (2016:244) indicate that the mechanisms that play a role in structural innovations rely on and result in maximal “explicitness of form-meaning relations” in both Learner and L2 Englishes. Similarly, Laporte (2012) reports the same patterns with regard to the use of *to*-infinitives in causative constructions (e.g. *to make someone to laugh*) in varieties of English as a Second Language (ESL) and varieties of English as a Foreign Language (EFL). Steger and Schneider (2012) observe an increase in the text frequency of overt complementizers in L2 varieties of English; Nesselhauf (2009) and Gilquin (2015) report usage of semantically redundant particles in (phrasal/prepositional) verbs in World Englishes (e.g. *enter into*) (see also Callies 2016:246); and Mesthrie (2006) discusses “anti-deletion” in some L2 varieties, a tendency to use explicit markers where speakers of native varieties would omit them.

4.2. Complexity versus simplicity

Closely linked to a variety's degree of analyticity or syntheticity is the structural parameter of simplicity versus complexity. Language complexity has been a popular topic in both in cross-linguistic typology and in dialect typology. The issue of cross-linguistic and intra-lingual complexity differentials has thereby increasingly gained attention, a view that diverges from the traditional view of twentieth-century structural linguists that all languages are equally complex (see Sampson 2009 for an overview; also Newmeyer and Preston 2014). The primary idea behind the equi-complexity hypothesis had been the assumption of a trade-off between the different subsystems of a language: complexity in one linguistic subsystem would be compensated by more simplicity in another subsystem. This trade-off hypothesis has been challenged at the beginning of the twenty-first century (see Gil 2008; Nichols 2009; Shosted 2006 whose empirical analyses give no indication of a trade-off; also see Sinnemäki 2014).

The discussion around this hypothesis is relevant for dialect typology. In a seminal paper challenging the equi-complexity hypothesis, McWhorter (2001) argued that creoles are less complex grammatically than their lexifier languages

by virtue of the fact that they were born as pidgins, and thus stripped of almost all features unnecessary to communication, and since then have not existed as natural languages for a long enough time for diachronic drift to create the weight of "ornament" that encrusts older languages. (McWhorter 2001:125)

And there is now an emerging consensus that language complexity is indeed variable (Miestamo, Sinnemäki and Karlsson 2008; Sampson, Gil and Trudgill 2009; see the papers in e.g. Kortmann and Szmrecsanyi 2012). Among other things, scholarship has sought to link observable complexity levels to language variation and change, for the sake of understanding simplification or complexification processes and the impact of sociolinguistic factors on these processes. The latter aspect has been of special interest: for instance, Trudgill (2001:372) links complexity to adult language learning when he states that "[a]dult language contact means adult language learning; and adult language learning means simplification, most obviously manifested in a loss of redundancy and irregularity and an increase in transparency". Childhood bilingualism on the other hand results in complexification (Trudgill 2011:42; see Section 2.2). Adult language learning and childhood bilingualism are intrinsically related to the outcome of language contact: low-contact varieties seem to exhibit more complexity than those communities that are, or have been, subject to intense contact with other languages or dialects (Trudgill 2011).

Linguists have proposed various measures according to which the complexity of *langue* or *parole* can be gauged. Most generally speaking, complexity measures can be dichotomized as follows (Miestamo 2008):

- *Global complexity measure versus local complexity measures*: global complexity quantifies the complexity of an entire language/dialect. Local complexity assesses the complexity of a domain-specific linguistic subsystem such as syntax or phonology.
- *Relative complexity measures versus absolute complexity measures*: measures of relative complexity gauge subjective, user-oriented complexity (related to processing and learning). Absolute complexity gauges objective, theory-oriented complexity by counting parts of the system such as the number of phonemes in a language.

A more fine-grained and detailed categorization could further distinguish redundancy-induced complexity measures, also called "ornamental complexity" where the amount of redundant linguistic material is counted towards a language's degree of complexity or absolute-quantitative complexity measures, where more material (larger marker inventories etc.) equals "more complex".

The bulk of complexity-oriented dialect-typological research is directed towards the investigation of differences in structural complexity between native and non-native varieties. Comparing indigenized L2 varieties and British English, Schneider (2015), for instance, highlights that future marker choice in Ghanaian English, an indigenized L2 variety, is less constrained by probabilistic factors than it is in British English. Assessing differences in the relative clause system between the same two varieties, Huber (2012) concludes that structural nativization in Ghanaian English has led there to an equally complex relativizer system in that the stochastic factors that impact relativizer choice are less or more important than in British English but they are overall the same. Steger and Schneider (2012) compare variable patterns in complement clause constructions between native and non-native varieties of English to explore the degree of iconicity and isomorphism in these new varieties. Iconicity refers to the relationship between form and meaning of a linguistic sign while isomorphism implies a one-to-one relationship between form and meaning. The authors define complexity as a function of iconicity, that is, iconicity effects lead to increased transparency of grammatical encoding and thus increased simplicity. Based on their findings, they argue that iconic constructions are more popular in non-native than in native varieties, and conclude that the cognitive principles at play during SLA lead to increased simplicity in second language varieties of English (Steger and Schneider 2012:187). Adopting a bird's eye perspective, Kortmann and Szmrecsanyi (2009) explore complexity patterns in the morphosyntax survey that accompanies the *Handbook of Varieties of English* (Kortmann et al. 2004). They classify the 76 morphosyntactic features covered in the handbook into (1) "simplifying features", that is, features or structures that simplify usage or the system vis-à-vis the standard English (an example would be leveling phenomena, such as leveling of preterite and past participle verb forms); and (2) "L2-simple features", i.e. features that are known to recur in interlanguage varieties, such as resumptive relative pronouns of the type *This is the house which I painted it yesterday* (see e.g. Hyltenstam 1984). Kortmann and Szmrecsanyi then quantify differences between the varieties sampled in the handbook on the basis of the number of simplifying and L2-simple features attested; the distributional pattern is visually depicted in Figure 7. The x-axis, labeled "Rule simplicity", plots the number of simplifying features; the y-axis plots the number of L2-simple features. Hence, varieties plotted in the upper-right corner attest a comparatively large amount of simplifying and L2-simple features; varieties plotted in the lower-left corner attest fewer simplifying and L2-simple features. Two observations can be highlighted: First, L2 simplicity predicts rule simplicity and vice versa. Second, English-based pidgin and creole languages are distinguished from other varieties: they cluster in the upper right-hand quadrant, indicating that they contain more simplifying and L2-simple features than other varieties. Surprisingly, indigenized L2 varieties are not really set apart from L1 varieties of English (see Kortmann & Szmrecsanyi 2009:276 for more discussion).

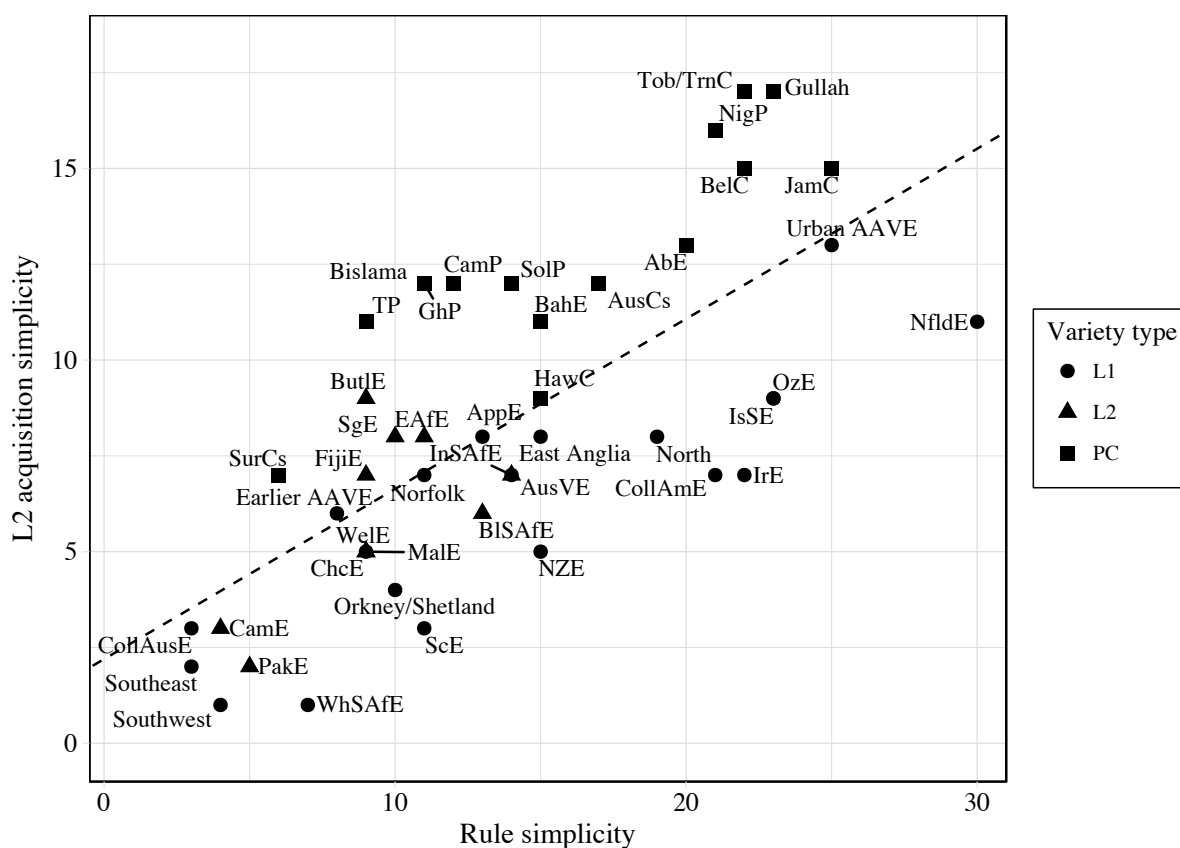


Figure 7. L2-simplicity by rule simplicity. The dotted trend line represents linear estimate of the relationship. (adapted from Kortmann and Szmrecsanyi 2009:Diagram 1)

5. Concluding remarks

Dialect typology is situated at the intersection between linguistic typology, a research field that is concerned with classifying human languages and with identifying structural similarities, and dialectology, which as a field is concerned with vernacular and regionally restricted and/or distinctive forms of language. Approaches that could not be addressed in this chapter but that are nonetheless relevant or at least neighboring to dialect typology include the extent to which variation patterns in particular varieties are bound by other factors not mentioned here (for instance, the degree of urbanization; see Donoso and Sanchez 2017); the potentially differential power of prescriptivism in different varieties (Hinrichs, Szmrecsanyi and Bohmann 2015); and work that generates typologies by considering attitudes and transnational importance (Mair 2013). Questions that have also not been addressed by the current chapter, but which constitute fascinating topics for future investigations include the extent to which differences and similarities between varieties on one grammatical level overlap with other linguistic levels (but see Glaser 2013:206; Scherrer and Stoeckle 2016), the extent to which semantic considerations might play a role in the characterization of dialectal differences, including cultural conceptualisations, and the degree to which these methods could lend support to a diachronic investigation of dialect diffusion. In this spirit, we would like to encourage research that adopts a more systematic approach to investigating the extent to which, for instance, the degree of urbanization – an under-researched factor that has been shown to be relevant in Spanish dialectology (e.g. Donoso & Sanchez 2017) – triggers dialectal divergence in varieties of, say, English or French. On more methodical grounds, we believe that the research community needs to abandon its traditional reliance on single-feature studies in favor of

methodologies that reveal large-scale patterns of variation based on multiple variables on multiple linguistic levels.

List of abbreviations of varieties

E/R/UAAVE	Earlier/Rural/Urban African American Vernacular English
AbE/AborE	(Australian) Aboriginal English
AppE	Appalachian English
AusCs	Australian Creoles
AusE	Australian English
AusVE	Australian Vernacular English
BahE	Bahamian English
BelC	Belizean Creole
BISAFé	Black South African English
ButIE	Butler English
CamP/E	Cameroon Pidgin/English
ChcE	Chicano English
ChIE	Channel Island English
CollAmE	Colloquial American English
CollAusE	Colloquial Australian English
CollSgE	Colloquial Singapore English
EA	East Anglian English
EAFé	East African English
FijiE	Fiji English
FlKE	Falkland English
GhP/E	Ghanaian Pidgin/English
HawC/E	Hawaiian Creole / English
InSAfE	Indian South African English
IrE	Irish English
IsSE/SEAmE	Isolated South Eastern American English
JamC/E	Jamaican Creole / English
LibSE	Liberian Settler English
MalE	Malaysian English
ManxE	Manx English
NfldE	Newfoundland English
NigP/E	Nigerian Pidgin / English
North	English dialects in the North of England
NZE	New Zealand English
O&SE	Orkney & Shetland English
OzE	Ozarks English
PacP	Pidgins of the Southwest Pacific (Tok Pisin, SolP, Bislama)
PakE	Pakistani English
ScE	Scottish English, Scots
SE	English dialects in the South-East of England
SEAmE / IsSE	South-Eastern US enclave dialects
SgE	Singapore English
SolP	Solomon Islands Pidgin
StHE	St. Helena English
SurC(s)	Suriname Creoles

SW	English dialects in the South-West of England
TdCE	Tristan da Cunha English
TobC	Tobagonian Creole
TrnC	Trinidadian Creole
TP	Tok Pisin, New Guinea Pidgin, Neomelanesian
Tob.TrnC	Creoles of Trinidad & Tobago
WeE	Welsh English
WhSAfE	White South African English
WhZimE	White Zimbabwean English

References

- Aikhenvald, Alexandra Y. & Robert M. W. Dixon (eds.). 2001. *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics*. Oxford: Oxford University Press.
- Anderwald, Lieselotte. 2003. *Negation in Non-Standard British English: Gaps, Regularizations and Asymmetries*. (Routledge Studies in Germanic Linguistics). London, New York: Routledge.
- Bickel, Balthasar. 2017. Areas and Universals. In Raymond Hickey (ed.), *The Cambridge Handbook of Areal Linguistics*, 40–54. Cambridge: Cambridge University Press.
- Callies, Marcus. 2016. Towards a process-oriented approach to comparing EFL and ESL varieties: A corpus-study of lexical innovations. *International Journal of Learner Corpus Research* 2(2). 229–250.
- Chambers, J. K. 2004. Dynamic typology and vernacular universals. In Bernd Kortmann (ed.), *Dialectology meets typology: Dialect grammar from a cross-linguistic perspective*, 127–145. Berlin/New York: Mouton de Gruyter.
- Crystal, David. 2004. *The Cambridge Encyclopedia of the English Language*. Second Edition. Cambridge: Cambridge University Press.
- Donoso, Gonzalo & David Sanchez. 2017. Dialectometric analysis of language variation in Twitter. *ArXiv e-prints*.
- Dryer, Matthew S. & Martin Haspelmath (eds.). 2013. *WALS Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/>.
- Dunn, Michael, Stephen C. Levinson, Eva Lindström, Ger Reesink & Angela Terrill. 2008. Structural Phylogeny in Historical Linguistics: Methodological Explorations Applied in Island Melanesia. *Language* 84(4). 710–759.
- Gil, David. 2008. How complex are isolating languages? In Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds.), *Language Complexity: Typology, Contact, Change*, 109–131. Amsterdam, Philadelphia: John Benjamins.
- Gilquin, Gaëtanelle. 2015. At the interface of contact linguistics and second language acquisition research: New Englishes and Learner Englishes compared. *English World-Wide* 36(1). 90–123.
- Glaser, Elvira. 2013. Area formation in morphosyntax. In Peter Auer, Martin Hilpert, Anja Stukenbrock & Benedikt Szmrecsanyi (eds.), *Space in language and linguistics: Geographical, interactional, and cognitive perspectives*, 195–221. (Linguae & Litterae 24). Berlin / Boston: Mouton de Gruyter.
- Gooskens, Charlotte. 2004. Norwegian dialect distances geographically explained. In Britt-Louise Gunnarson, Lena Bergström, Gerd Eklund, Staffan Fridella, Lise H. Hansen, Angela Karstadt, Bengt Nordberg, Eva Sundgren & Mats Thelander (eds.), *Language variation in Europe. Papers from the Second International Conference on Language Variation in Europe ICLAVE*, vol. 2, 195–206. Uppsala.
- Grafmiller, Jason & Benedikt Szmrecsanyi. in press. Mapping out particle placement in varieties of English across the globe. *Language Variation and Change*.
- Greenberg, Joseph H. 1960. A quantitative approach to the morphological typology of language. *International Journal of American Linguistics* 26(3). 178–194.
- Greenberg, Joseph H. 1963. *The Languages of Africa*. Bloomington: Indiana University.

- Greenberg, Joseph H. 1966. *Language Universals, With Special Reference to Feature Hierarchies*. The Hague: Mouton.
- Hengeveld, Kees & Sterre Leufkens. 2018. Transparent and non-transparent languages. *Folia Linguistica* 52(1). 139–175.
- Hinrichs, Lars, Benedikt Szmrecsanyi & Axel Bohmann. 2015a. Which-hunting and the Standard English relative clause. *Language* 91(4). 806–836.
- Huber, Magnus. 2012. Syntactic and variational complexity in British and Ghanaian English: Relative clause formation in the written parts of the International Corpus of English. In Bernd Kortmann & Benedikt Szmrecsanyi (eds.), *Linguistic Complexity: Second Language Acquisition, Indigenization, Contact*, 218–242. Berlin: Mouton De Gruyter.
- Hyltenstam, Kenneth. 1984. The use of typological markedness conditions as predictors in second language acquisition: The case of pronominal copies in relative clauses. In R. Andersen (ed.), *Second Languages*, 39–58. Rowley, MA: Newbury.
- Jeszszky, Péter, Philipp Stoeckle, Elvira Glaser & Robert Weibel. 2017. Exploring global and local patterns in the correlation of geographic distances and morphosyntactic variation in Swiss German. *Journal of Linguistic Geography* 5(2). 86–108. doi:10.1017/jlg.2017.5.
- Kachru, Braj B. (ed.). 1992. *The Other tongue: English across cultures*. 2nd ed. (English in the Global Context). Urbana: University of Illinois Press.
- Kortmann, Bernd (ed.). 2004. *Dialectology Meets Typology: Dialect Grammar from a Cross-Linguistic Perspective*. Berlin, New York: Mouton de Gruyter.
- Kortmann, Bernd & Kerstin Lunkenheimer. 2012. *The Mouton World Atlas of Variation in English*. Berlin: de Gruyter Mouton.
- Kortmann, Bernd & Kerstin Lunkenheimer (eds.). 2013. *eWAVE*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://ewave-atlas.org/>.
- Kortmann, Bernd, Edgar Schneider, Kate Burridge, Raj Mesthrie & Clive Upton (eds.). 2004. *A Handbook of Varieties of English*. Berlin: Mouton de Gruyter.
- Kortmann, Bernd & Benedikt Szmrecsanyi. 2004. Global synopsis: Morphological and syntactic variation in English. In Bernd Kortmann, Edgar W. Schneider, Kate Burridge, Rajend Mesthrie & Clive Upton (eds.), *A handbook of varieties of English*, vol. 2: Morphology and syntax, 1142–1202. Berlin/New York: Mouton de Gruyter.
- Kortmann, Bernd & Benedikt Szmrecsanyi. 2009. World Englishes between simplification and complexification. In Lucia Siebers & Thomas Hoffmann (eds.), *World Englishes -- Problems, Properties and Prospects: selected papers from the 13th IAWE conference*, 265–285. Amsterdam: John Benjamins.
- Kortmann, Bernd & Benedikt Szmrecsanyi. 2011. Parameters of morphosyntactic variation in World Englishes: Prospects and limitations of searching for universals. In Peter Siemund (ed.), *Linguistic Universals and Language Variation*, 264–290. Berlin / New York: Mouton de Gruyter.
- Kortmann, Bernd & Benedikt Szmrecsanyi. 2012. *Linguistic complexity: Second language acquisition, indigenization, contact*. Berlin, Boston: Mouton de Gruyter.
- Kruskal, Joseph B & Myron Wish. 1978. *Multidimensional Scaling*. (Quantitative Applications in the Social Sciences). Newbury Park, London, New Delhi: Sage Publications.
- Kusters, Wouter. 2003. Linguistic complexity: the influence of social change on verbal inflection. Leiden: Universiteit Leiden PhD.
- Laporte, Samantha. 2012. Mind the gap! Bridge between World Englishes and Learner Englishes in the making. *English Text Construction* 5(2). 265–292. doi:10.1075/etc.5.2.05lap.
- Lê, Sébastien, Julie Josse & François Husson. 2008. FactoMineR: an R package for multivariate analysis. *Journal of Statistical Software* 25(1). 1–18.
- Leufkens, Sterre. 2013. The transparency of creoles. *Journal of Pidgin and Creole Languages* 28(2). 323–362.
- Levshina, Natalia. 2015. *How to do linguistics with R: data exploration and statistical analysis*. Amsterdam ; Philadelphia: John Benjamins Publishing Company.
- Lipski, John M. 2008. *Afro-Bolivian Spanish*. Madrid: Iberoamericana.

- Mair, Christian. 2013. The World System of Englishes: Accounting for the transnational importance of mobile and mediated vernaculars. *English World-Wide* 34(3). 253–278.
doi:10.1075/eww.34.3.01mai.
- Matras, Yaron. 2009. *Language contact*. Cambridge: Cambridge University Press.
- McMahon, April, Paul Heggarty, Robert McMahon & Warren Maguire. 2007. The sound patterns of Englishes: representing phonetic similarity. *English Language and Linguistics* 11(1). 113–142.
- McWhorter, John. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* 5. 125–166.
- Mesthrie, Rajend. 2006. Anti-deletions in an L2 grammar: A study of Black South African English mesolect. *English World-Wide* 27(2). 111–145.
- Michaelis, Susanne Maria, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.). 2013. *APiCS Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://apics-online.info/>.
- Miestamo, Matti. 2008. Grammatical complexity in a cross-linguistic perspective. In Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds.), *Language Complexity: Typology, Contact, Change*, 23–41. Amsterdam: John Benjamins.
- Miestamo, Matti, Kaius Sinnemäki & Fred Karlsson (eds.). 2008. *Language complexity: Typology, contact, change*. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Montemagni, Simonetta. 2008. The space of Tuscan dialectal variation: A correlation study. *International Journal of Humanities and Arts Computing* 2. 135–152.
- Murelli, Adriano & Bernd Kortmann. 2011. 28 Non-standard varieties in the areal typology of Europe. In Bernd Kortmann & Johan van der Auwera (eds.), *The Languages and Linguistics of Europe*. Berlin, Boston: DE GRUYTER MOUTON.
<http://www.degruyter.com/view/books/9783110220261/9783110220261.525/9783110220261.525.xml> (4 October, 2016).
- Nerbonne, John. 2010. Measuring the diffusion of linguistic change. *Philosophical transactions of the Royal Society of London, Series B, Biological Sciences* 365(1559). 3821–3828.
- Nerbonne, John & Peter Kleiweg. 2007. Toward a Dialectological Yardstick. *Journal of Quantitative Linguistics* 14(2). 148–166.
- Nesselhauf, Nadja. 2009. Co-selection phenomena across New Englishes. *English World-Wide* 30(1). 1–26.
- Newmeyer, Frederick J & Laurel B. Preston (eds.). 2014. *Measuring grammatical complexity*. Oxford: Oxford University Press.
- Nichols, Johanna. 2009. Linguistic complexity: a comprehension definition and survey. In Geoffrey Sampson, David Gil & Peter Trudgill (eds.), *Language complexity as an evolving variable*, 64–79. Oxford: Oxford University Press.
- Nichols, Johanna. 2013. The vertical archipelago: Adding the third dimension to linguistic geography. In Peter Auer, Martin Hilpert, Anja Stukenbrock & Benedikt Szmrecsanyi (eds.), *Space in Language and Linguistics: Geographical, interactional, and cognitive perspectives*, 38–60. Berlin / Boston: De Gruyter.
- Perez, Danae M. forthcoming. Reassessing the family of Afro-Iberian contact varieties: From Príncipe to Potosí.
- Perez, Danae M. 2015. Traces of Portuguese in Afro-Yungueño Spanish? *Journal of Pidgin and Creole Languages* 30. 307–343.
- Poplack, Shana & Sali A. Tagliamonte. 2001. *African American English in the diaspora*. Oxford: Blackwell.
- Röthlisberger, Melanie. 2018. Regional variation in probabilistic grammars: A multifactorial study of the English dative alternation. Leuven: KU Leuven PhD dissertation.
- Sampson, Geoffrey. 2009. A linguistic axiom challenged. In Geoffrey Sampson, David Gil & Peter Trudgill (eds.), *Language complexity as an evolving variable*, 1–18. Oxford: Oxford University Press.
- Sampson, Geoffrey, David Gil & Peter Trudgill. 2009. *Language Complexity as an Evolving Variable*. (Ed.) Geoffrey Sampson, David Gil & Peter Trudgill. Oxford: Oxford University Press.

- Sapir, Edward. 1921. *Language: An introduction to the study of speech*. New York: Harcourt, Brace and Company.
- Scherrer, Yves & Philipp Stoeckle. 2016. A quantitative approach to Swiss German - Dialectometric analyses and comparison of linguistic levels. *Dialectologia et Geolinguistica* 24. 92–125.
- Schlegel, August Wilhelm von. 1818. *Observations sur la langue et la littérature provençales*. Paris.
- Schneider, Agnes. 2015. Aspect and Modality in Ghanaian English: A Corpus-based Study of the Progressive and the Modal WILL. University of Freiburg PhD dissertation.
- Schneider, Edgar. 2004. Global synopsis: phonetic and phonological variation in English world-wide. In Bernd Kortmann, Edgar Schneider, K. Burridge, R. Mesthrie & C. Upton (eds.), *A Handbook of Varieties of English*, vol. 1, 1111–1137. Berlin/New York: Mouton de Gruyter.
- Schreier, Daniel. 2016. A true split? Typological and sociolinguistic considerations on contact intensity effects. In Raffaella Baechler & Guido Seiler (eds.), *Complexity, Isolation, and Variation*, 139–157. (Linguae & Litterae 57). Berlin, Boston: Mouton de Gruyter.
- Sessarego, Sandro. 2013. On the non-creole basis for Afro-Bolivian Spanish. *Journal of Pidgin and Creole Languages* 28. 363–407.
- Shackleton, Robert G. Jr. 2007. Phonetic variation in the traditional English dialects: a computational analysis. *Journal of English Linguistics* 35(1). 30–102.
- Shosted, Ryan. 2006. Correlating complexity: A typological approach. *Linguistic Typology* 10. 1–40.
- Siegel, Jeff, Benedikt Szmrecsanyi & Bernd Kortmann. 2014. Measuring analyticity and syntheticity in creoles. *Journal of Pidgin and Creole Languages* 29(1). 49–85. doi:10.1075/jpcl.29.1.02sie.
- Siemund, Peter. 2008. Language contact: Constraints and common paths of contact-induced language change. In Peter Siemund & Noemi Kintana (eds.), *Language Contact and Contact Languages*, 3–11. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Siemund, Peter & Noemi Kintana (eds.). 2008. *Language contact and contact languages*. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Sinnemäki, Kaius. 2014. Complexity trade-offs: A case study. In Frederick J. Newmeyer & Laurel B. Preston (eds.), *Measuring Grammatical Complexity*, 179–201. Oxford: Oxford University Press.
- Spruit, Marco René. 2008. *Quantitative perspectives on syntactic variation*. Utrecht: LOT.
- Spruit, Marco René, Wilbert Heeringa & John Nerbonne. 2009. Associations among linguistic levels. *Lingua* 119(11). 1624–1642.
- Steger, Maria & Edgar W Schneider. 2012. Complexity as a function of iconicity: The case of complement clause constructions in New Englishes. In Bernd Kortmann & Benedikt Szmrecsanyi (eds.), *Linguistic Complexity: Second Language Acquisition, Indigenization, Contact*. Berlin: de Gruyter.
- Szmrecsanyi, Benedikt. 2009. Typological parameters of intralingual variability: Grammatical analyticity versus syntheticity in varieties of English. *Language Variation and Change* 21(03). 319–353. doi:10.1017/S0954394509990123.
- Szmrecsanyi, Benedikt. 2012a. Typological profile: L1 varieties. In Bernd Kortmann & Kerstin Lunkenheimer (eds.), *The Mouton World Atlas of Variation in English*, 826–843. Berlin/Boston: Mouton de Gruyter.
- Szmrecsanyi, Benedikt. 2012b. Geography is overrated. In Sandra Hansen, Christian Schwarz, Philipp Stoeckle & Tobias Streck (eds.), *Dialectological and folk dialectological concepts of space: Current methods and perspectives in sociolinguistic research on dialect change*, 215–231. Berlin / Boston: De Gruyter.
- Szmrecsanyi, Benedikt. 2017. “Featurometry.” In Martijn Wieling, Martin Kroon, Gertjan van Noord & Gosse Bouma (eds.), *From semantics to dialectometry: Festschrift in honor of John Nerbonne*, 345–353. College Publications.
- Szmrecsanyi, Benedikt & Bernd Kortmann. 2009a. Vernacular universals and angloversals in a typological perspective. In Markku Filppula, Juhani Klemola & Heli Paulasto (eds.), *Vernacular Universals and Language Contacts: Evidence from Varieties of English and Beyond*, 33–53. London, New York: Routledge.

- Szmrecsanyi, Benedikt & Bernd Kortmann. 2009b. The morphosyntax of varieties of English worldwide: a quantitative perspective. *Lingua* 119(11). 1643–1663.
- Szmrecsanyi, Benedikt & Bernd Kortmann. 2009c. Between simplification and complexification: non-standard varieties of English around the world. In Geoffrey Sampson, David Gil & Peter Trudgill (eds.), *Language Complexity as an Evolving Variable*, 64–79. Oxford: Oxford University Press.
- Szmrecsanyi, Benedikt & Melanie Röthlisberger. to appear. World Englishes from the perspective of dialect typology. In Daniel Schreier, Marianne Hundt & Edgar W. Schneider (eds.), *The Cambridge handbook of World Englishes*. Cambridge: Cambridge University Press.
- Szmrecsanyi, Benedikt & Bernhard Wälchli (eds.). 2014. *Aggregating dialectology, typology, and register analysis: linguistic variation in text and speech*. (Lingua & Litterae 28). Berlin: Walter de Gruyter.
- Tagliamonte, Sali A. 2002. Comparative sociolinguistics. In J. K. Chambers, Peter Trudgill & Natalie Schilling-Estes (eds.), *The handbook of language variation and change*, 729–763. Malden/Oxford: Blackwell.
- Trudgill, Peter. 1996. Dialect typology: isolation, social network and phonological structure. In Gregory R. Guy, Crawford Feagin, Deborah Schiffrin & John Baugh (eds.), *Towards a Social Science of Language: Papers in honor of William Labov. Volume 1: Variation and change in language and society*, 3. Amsterdam: John Benjamins Publishing Company.
<https://benjamins.com/catalog/cilt.127.04tru> (7 September, 2016).
- Trudgill, Peter. 2001. Contact and simplification: Historical baggage and directionality in linguistic change. *Linguistic Typology* 5. 371–374.
- Trudgill, Peter. 2004. Linguistic and Social Typology: The Austronesian migrations and phoneme inventories. *Linguistic Typology* 8. 305–320.
- Trudgill, Peter. 2006. Dialect mixture versus monogenesis in colonial varieties: The inevitability of Canadian English? *Canadian Journal of Linguistics-Revue Canadienne De Linguistique* 51. 265–286. doi:10.1353/cjl.2008.0002.
- Trudgill, Peter. 2008. The role of dialect contact in the formation of Englishes. In Miriam A. Locher & Jürg Strässler (eds.), *Standards and Norms in the English Language*, 69–83. Berlin: Mouton de Gruyter.
- Trudgill, Peter. 2009a. Sociolinguistic typology and complexification. In Geoffrey Sampson, David Gil & Peter Trudgill (eds.), *Language Complexity as an Evolving Variable*, 98–109. Oxford: Oxford University Press.
- Trudgill, Peter. 2009b. Vernacular Universals and the Sociolinguistic Typology of English Dialects. In Markku Filppula, Juhani Klemola & Heli Paulasto (eds.), *Vernacular Universals and Language Contact: Evidence from Varieties of English and Beyond*, 304–322. (309). London: Routledge.
- Trudgill, Peter. 2009c. Vernacular universals and the sociolinguistic typology of English dialects. In Markku Filppula, Juhani Klemola & Heli Paulasto (eds.), *Vernacular Universals and Language Contacts: Evidence from Varieties of English and Beyond*, 302–329. London: Routledge.
- Trudgill, Peter. 2010. Contact and Sociolinguistic Typology. In Raymond Hickey (ed.), *The handbook of language contact*, 299–319. Malden: Wiley-Blackwell.
- Trudgill, Peter. 2011. *Sociolinguistic typology: social determinants of linguistic complexity*. Oxford, New York: Oxford University Press.
- Trudgill, Peter, Elizabeth Gordon, Gillian Lewis & Margaret MacLagan. 2000. Determinism in New-Dialect Formation and the Genesis of New Zealand English. *Journal of Linguistics* 36(2). 299–318.