

2 Vernacular Universals and Angloverals in a Typological Perspective

Benedikt Szmrecsanyi and Bernd Kortmann

1. INTRODUCTION

In this study, we endeavor to take a typological perspective on language-internal variation in English. Our inquiry will be based on the largest comparative survey to date of entire grammatical subsystems of varieties of English worldwide (Kortmann and Szmrecsanyi 2004a, 2004b). A catalogue of 76 morphosyntactic features taken from the 11 core areas of English morphosyntax will be investigated for 46 (groups of) nonstandard varieties of English around the world. Our analytical point of departure is the notion that there are different reasons that languages, or varieties of a given language, should exhibit the same linguistic features. Such features may fall into any one of the following categories:

- (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*).

With areoversals constituting the level of generalization below which orthodox linguistic typology normally does not stray, the present chapter

focuses on (iv)–(vii) in varieties of English worldwide. Taking the notion of ‘vernacular universals’ (cf., for instance, Chambers 2004) as its starting point, the present study embarks on an in-depth exploration of candidate (uni)versals and some implicational tendencies in different types and areal groups of the world’s Englishes. We also utilize principal component analysis to explore large-scale parameters of variance, which are possibly recurrent cross-linguistically, in our global survey of varieties of English. On a methodological note, then, this study is in keeping with the Freiburg program of combining functional Greenbergian typology with dialectology, which argues, in a nutshell, that the observable patterns of cross-dialectal and cross-varietal variation can be best analyzed and interpreted in terms of the same methodological and interpretational apparatus familiar from the typological study of large-scale cross-linguistic variation (see Kortmann 2004b for a collection of papers in this spirit).

2. DATA: THE HANDBOOK OF VARIETIES OF ENGLISH

This study rests empirically on the *Handbook of Varieties of English* (Kortmann et al. 2004), the most comprehensive reference work to date on the phonology (Volume 1) and morphosyntax (Volume 2) of varieties of English around the globe. The *Handbook* contains survey articles by about 100 specialists, covering some 60 almost exclusively nonstandard varieties or groups of varieties: native vernaculars (henceforth: L1 varieties), which include all main national standard varieties, distinctive ethnic, regional, and social varieties, e.g. African American Vernacular English; English-based pidgins and creoles, such as Tok Pisin; and the major English as a Second Language (henceforth: L2 varieties) varieties, e.g. Malaysian English. Crucially, all of the varieties surveyed are spoken ones.

What will take center stage in the present study is the morphosyntactic survey of the multimedia reference tool, available on CD-ROM and online (<http://www.mouton-online.com>), accompanying the *Handbook* (cf. Kortmann and Szmrecsanyi 2004). This first-ever comprehensive survey of nonstandard English morphosyntax worked in a very simple way: We compiled a catalogue of 76 features—essentially, the ‘usual suspects’ in previous dialectological, variationist, and creolist research—and sent out this catalogue to the authors of the chapters in the morphosyntax volume of the *Handbook*. For each of these 76 features, the contributors were asked to specify into which of the following three categories the relevant feature falls:

- 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



Figure 2.1 Varieties sampled in the morphosyntactic database of the *Handbook*.

Kortmann and Szmrecsanyi (2004: 1142–5) discuss the survey procedure, the varieties of English sampled, as well as the advantages and drawbacks of the method in considerable detail. Suffice it to say here that the 40 *Handbook* authors provided us with data on 46 nonstandard varieties of English. **Figure 2.1** visualizes the geographical distribution of varieties of English in the survey. As can be seen, all seven anglophone world regions (British Isles, America, Caribbean, Australia, Pacific, Asia, and Africa), as well as a fair mix of L1 varieties, L2 varieties, and pidgins/creoles are represented in the survey.

The features are numbered from 1 to 76 (see the appendix for the feature catalogue in its entirety) and include all major phenomena discussed in previous survey articles on grammatical properties of (individual groups of) nonstandard varieties of English, with a slight bias towards features observed in L1 varieties. They cover 11 broad areas of morphosyntax: pronouns, the noun phrase, tense and aspect, modal verbs, verb morphology, adverbs, negation, agreement, relativization, complementation, and discourse organization and word order.

3. VERNACULAR UNIVERSALS OR VERNACULAR ANGLOVERSALS?

We have drawn on the *Handbook*'s database elsewhere to offer an extended empirical discussion of (i) unrestricted angloversals, i.e. the most frequent features in varieties of English across the board (Kortmann and Szmrecsanyi 2004: 1153–60); (ii) unrestricted areoversals (Kortmann and Szmrecsanyi 2004: 1160–84); and (iii) unrestricted varioversals, i.e. top features in New

Englishes (Kortmann and Szmrecsanyi 2004: 1184–94). Here, we seek to explore in more depth the socio-dialectological notion of ‘vernacular universals’ or ‘vernacular roots’. According to Jack Chambers, these universals comprise “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). Crucially, the putative ubiquity of such features is argued to be unlikely due to sociolinguistic diffusion, which is why they must be “primitive features of vernacular dialects” (Chambers 2003: 243)—that is, unlearned and thus innate. Chambers (2004: 129) lists the following four morphosyntactic candidates for vernacular universals (indicated in square brackets are the features in our 76-features catalogue which correspond most closely to the four morphosyntactic processes named by Chambers):

- (1) conjugation regularization, or levelling of irregular verb forms: *John seen the eclipse, Mary heard the good news* [36–39];
- (2) default singulars, or subject–verb nonconcord: *They was the last ones* [55, 59; marginally 53 and 54];
- (3) *multiple negation, or negative concord* [44];
- (4) copula absence, or copula deletion: *She smart, We going as soon as possible* [57; possibly 58, 73].

So, how frequent are these features in vernacular varieties of English sampled in our database? **Figure 2.2**, which plots the percentages of varieties where they are attested (as either pervasive or existing, but rare), provides details. Certainly, these features are rather frequent: Auxiliary deletion in *wh*-questions [73] recurs in 78% of the varieties in the survey, multiple negation [44] is attested in 76%, and regularization of irregular verb paradigms [36] in 70%. Crucially, Figure 2.2 also suggests that in Englishes spoken in

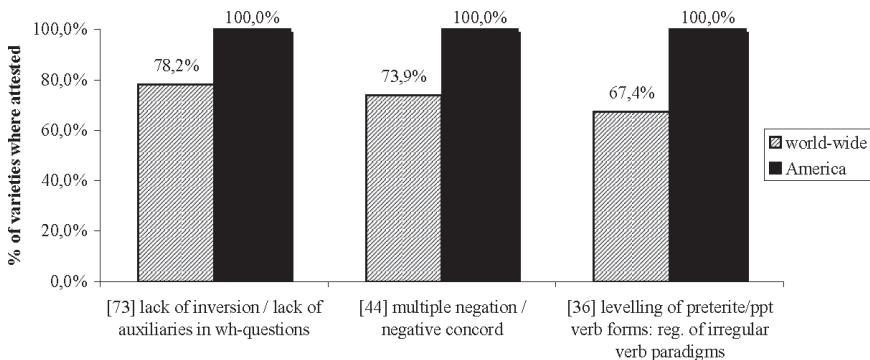


Figure 2.2 Frequencies of some candidates for vernacular angloversals in the *Handbook's* database.

the Americas, these features are even more widespread; indeed, it is here that they are truly universal. Yet on a global scale, none of the features is attested in more than 80% of varieties of English worldwide; thus, frequent they are, but universal they are not. What is more, there are features in our survey which are considerably more pervasive globally (although note that no feature in our 76-feature survey is actually attested in each and every variety sampled). This is why we would like to suggest that these features—conjugation regularization, multiple negation, and copula/auxiliary absence—be more appropriately considered areoversals precisely because of their preponderance in vernacular Englishes spoken in America.

Along these lines, we believe that some terminological caution might be warranted in the use of the term ‘universal’. Notice here that this would be true even if we found a nonstandard feature attested in each and every variety of English in the world: According to most definitions, vernacular *universals* should have counterparts in the vernaculars of other languages as they are “the outgrowths of [. . .] rules and representations in the bio-program” (Chambers 2004: 129). Yet, of the four candidates given, multiple negation is, at least at first glance, the only convincing one in a truly cross-linguistic perspective. What about copula absence? It turns out that in a cross-linguistic perspective, zero copulas are simply not that frequent. **Figure 2.3** reproduces Map 120 (Stassen 2005) of the *World Atlas of Language Structures* (Haspelmath et al. 2005) which plots languages where zero copula for predicate nominals (of the type *He __ a sailor*) is possible against languages where it is impossible. What emerges is that, in a significant majority (211: 175) of languages sampled by Stassen, pronominal zero copulas are impossible. If zero copulas were somehow part of the language faculty,

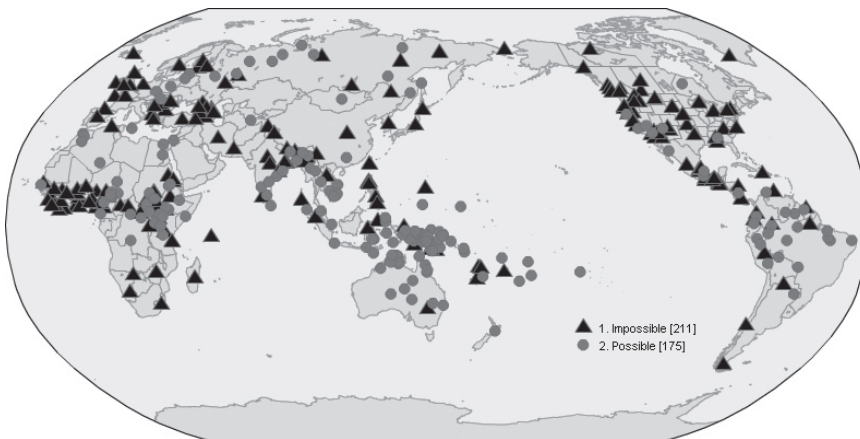


Figure 2.3 Zero copula for predicate nominals—the cross-linguistic perspective (Stassen 2005). Black triangles indicate languages where the phenomenon is impossible, and gray dots indicate languages where the phenomenon is possible. Sample size: 386 languages.

there should be more languages that have them—and this means that the phenomenon is, as a matter of fact, not such a promising candidate for a vernacular *universal*.¹

As for subject–verb nonconcord (or default singulars), we wish to highlight that a language needs structural means to display subject–verb concord (at least historically) to exhibit subject–verb nonconcord in any nonvacuous way. Yet many languages—for instance, highly isolating languages such as Vietnamese—do not show agreement at all; so vernacular Vietnamese showing nonconcord is an utterly unremarkable fact. Conversely, one encounters the problem that there are many languages (e.g. Hungarian) that have a lot more agreement than English; this is unexpected if the ‘bioprogram’ has a proclivity toward nonconcord. Along similar lines, a language must have irregular verb forms (again, at least historically) to exhibit conjugation regularization. Yet Turkish, for instance, is commonly considered not to have irregular verbs—how could it, then, display conjugation regularization? It might be contended that Turkish does not, in fact, refute the argument because Turkish can be seen as a maximally consistent implementation of the ‘conjugation regularization’ principle. But then again, if total consistency is theoretically possible, every single language with a good deal of irregular verbs (e.g. Italian) would play havoc with the bioprogram explanation.

In short, what we mean to suggest here is that features such as default singulars or conjugation regularization are things which are rather typical of languages that, like English, have some inflectional morphology but are in the process of getting rid of what remains. But what is happening in non-standard varieties of English and, possibly, languages belonging to the same morphological type as English does not necessarily apply to vernaculars of inflectional or agglutinating languages such as Finnish, Hungarian (which have a great deal of grammatical agreement), or Turkish. A feature such as default singulars is maybe better referred to as a vernacular *typoversal*—a feature, in other words, which is typical of vernaculars of inflecting languages. It is unlikely that the ‘language faculty’ would provide for special rules and representations applying to English-like vernaculars; default singulars and conjugation regularization are just too conditioned on the typological type of English to be cross-linguistically ‘universal’. At the same time, notice that it is not only loss of agreement or loss of redundancy that we can observe in vernaculars; individual vernaculars have and can indeed be shown to currently develop a more elaborate inflectional morphology or, for example, an agreement system than the standard variety has (cf. several studies in Barbiers et al. 2002; Kortmann 2004b). Given these observations, we believe that a candidate feature for a vernacular universal should, at a minimum, fulfill the following criteria:

- The candidate feature should be attested in a vast majority of a given language’s vernacular varieties.

- The candidate feature should not be patterned geographically or according to variety type (in the case of English: L1, L2, or pidgin/creole).
- For the sake of cross-linguistic validity, the candidate feature should not be tied to a given language's typological make-up (inflectional, agglutinating, etc.).²
- The candidate feature should be cross-linguistically attested in a significant number of the world's languages (especially among the world's many languages without a literary/philosophical tradition).

4. VARIETIES OF ENGLISH: IMPLICATIONAL TENDENCIES

We now turn to a number of correlational tendencies and implications between individual nonstandard features in the *Handbook's* survey. Analysts of cross-linguistic variation make a distinction between two types of correlational tendencies: 'biconditional implications', also known as 'equivalences' (e.g. 'if in a language the genitive follows the noun, then the complement follows the adposition, and vice versa'; Greenberg 1963), and 'one-way implications', also known as 'preferences' (e.g. 'if a language has a marked singular, it has also a marked plural, but not necessarily vice versa'; Greenberg 1966). Crucially, most analysts consider that the correlations need not be absolute or perfect, but can have exceptions.

4.1 Biconditional Implications

Let us begin by discussing some biconditional implications (or equivalences) exhibited in the *Handbook's* survey. To reiterate, biconditional implications are nonhierarchical correlations, such that varieties will attest just one of two features, but not the other. Out of $76 \times 75/2 = 2,850$ potential candidate pairings (thanks to the survey's coverage of 76 nonstandard features), there are no biconditional implications which hold, without exceptions, for *all* varieties in our sample. Admitting a limited number of counterexamples as acceptable, however, the following biconditional implications emerge as the front runners for biconditional-implicational angloverals:

- [45] *ain't* as the negated form of *be*—[46] *ain't* as the negated form of *have* (this biconditional implication holds in 94% of all varieties)
- [12] noncoordinated subject pronoun forms in object function—[13] noncoordinated object pronoun forms in subject function (89%)
- [23] habitual *do*—[27] *do* as a tense and aspect marker (89%)
- [63] relative particle *as*—[64] relative particle *at* (89%)

So, 94% of the varieties sampled either have both *ain't* as the negated form of *be* and *ain't* as the negated form of *have* or they have neither of

these things. This relationship—which is consonant with many previous dialect descriptions (cf. Anderwald 2003: 149–50)—can be neatly visualized in a tetrachoric table where unpredicted cells are shaded in gray:

ain't as negated form of *have*

		attested	not attested
<i>ain't</i> as negated form of <i>be</i>	attested	AusVE, . . .	(IrE, WelE)
	not attested	(NZE)	ScE, . . .

Australian Vernacular English is a typical variety in that it exhibits both *ain't* as the negated form of *be* and *ain't* as the negated form of *have*. Scottish English also conforms with the biconditional implication, in that it has neither of these features. New Zealand English, Irish English, and Welsh English, by contrast, are the three exceptions in our database to the statistical tendency in that only one but not the other of the uses of *ain't* is attested in them.

In a similar vein, noncoordinated subject pronoun forms in object function [12] and noncoordinated object pronoun forms in subject function [13] also tend to go together, as do habitual *do* [23] and *do* as a tense and aspect marker [27].³ Likewise, if a variety exhibits the relative particle *at* [64], it will have the relative particle *as* [63] and vice versa. In short, pairings such as these are best seen as feature bundles instead of pairs of independent features.

What is to be said about *biconditional-implicational varioversals*? Distinguishing among L1 varieties, L2 varieties, and pidgins and creoles, Table 2.1 provides a list of some perfect—i.e. exceptionless-biconditional implications. Thus, among L1 varieties (Table 2.1a), we find a correlation such that when in a variety past forms of irregular verbs can replace participle forms [38], unmarked forms [37] are also possible and vice versa. Further, if a L1 vernacular displays noncoordinated object pronoun forms in subject function [13], noncoordinated subject pronoun forms in object function [12] are also possible and vice versa. No L1 vernacular in our sample, then, displays just one of these things.

Among both L2 varieties and English-based pidgins and creoles (Table 2.1b–c), we observe that if a variety has one use of *ain't* (either for *be*, *have*, or as a generic negator), it will also have the other two uses of the form. This versatility of *ain't*, then, can be considered a biconditional-implicational universal of New Englishes. In L2 varieties, we also note, for instance, that if a variety has *no* as a preverbal negator [50], it will also display the past tense/anterior marker *been* [29] and vice versa. In English-based pidgins and creoles (Table 2.1c), lack of inversion in main clause *yes/no* questions [74] and lack of inversion/lack of auxiliaries in *wh*-questions [73] always occur in tandem, which is another way of saying that English-based pidgins and creoles either invert in questions or they don't.

Table 2.1 Some Perfect (i.e. Exceptionless) Biconditional-Implicational Varioversals

Feature 1	Feature 2
<i>a. L1 Varieties</i>	
[38] levelling of preterite/ppt verb forms: past replacing the participle	[37] levelling of preterite/ppt verb forms: unmarked forms
[13] noncoordinated object pronoun forms in subject function	[12] noncoordinated subject pronoun forms in object function
<i>b. L2 Varieties</i>	
[46] <i>ain't</i> as the negated form of <i>have</i>	[45/47] <i>ain't</i> as the negated form of <i>be/ain't</i> as generic negator before a main verb
[50] <i>no</i> as preverbal negator	[29] past tense/anterior marker <i>been</i>
<i>c. English-Based Pidgins and Creoles</i>	
[46] <i>ain't</i> as the negated form of <i>have</i>	[45/47] <i>ain't</i> as the negated form of <i>be/Ain't</i> as generic negator before a main verb
[74] lack of inversion in main clause <i>yes/no</i> -questions	[73] lack of inversion / lack of auxiliaries in <i>wh</i> -questions

Exhibited in the database: the occurrence/non-occurrence of feature 1 is conditioned on the occurrence/non-occurrence of feature 2, and vice versa. The designation of pair parts as ‘feature 1’ and ‘feature 2’, respectively, is arbitrary.

4.2 One-Way Implications

We now turn our attention to the one-way implication of the type ‘if A then B but not necessarily vice versa’. Overall, the varieties sampled in our survey exhibit the surprisingly large number of 247 perfect (i.e. 100% felicitous) bidirectional statistical correlations, out of $76 \times 75 = 5,700$ potential candidate pairings; at the 85% felicity level, there are no less than 721 such correlations. Needless to say, a good deal of these *prima facie* implications turn out, upon closer inspection, to be meaningless due to lack of statistical significance (cf. Cysouw 2003). Still, the database appears to contain a healthy number of substantial one-way implications. In an inevitably eclectic fashion, Table 2.2 lists some perfect and statistically significant one-way implications, indicating statistical robustness in the rightmost column.⁴

In Table 2.2a, we find a number of one-way implicational angloversals, i.e. one-way implications that hold for each and every variety of English in the sample. Thus, any variety that has *would* in *if*-clauses [31] also displays loosening of the sequence of tense rule [30], but not necessarily vice versa. The following tetrachoric table illustrates this (as before, unpredicted cells are shaded in gray):

Table 2.2 Some Perfect (i.e. Exceptionless) One-Way Implicational Angloverals (a.) and Areoversals (b.–d.)

<i>Feature 1</i>	<i>Feature 2</i>	<i>Fisher's Exact Test</i>
<i>a. All Varieties</i>		
[31] <i>would</i> in <i>if</i> clauses	[30] loosening of sequence of tense rule	$p < .001$
[47] <i>ain't</i> as generic negator before a main verb	[45/46] <i>ain't</i> as the negated form of <i>belhave</i>	$p < .005$
[63] relative particle <i>as</i>	[61] relative particle <i>what</i> / [66] gapping or zero-relativization in subject position	$p < .05$
<i>b. L1 Varieties</i>		
[27] <i>do</i> as a tense and aspect marker	[23] habitual <i>do</i>	$p < .005$
[45] <i>ain't</i> as the negated form of <i>be</i>	[44] multiple negation / negative concord	$p < .05$
[22] habitual <i>be</i>	[26] <i>be</i> as perfect auxiliary	$p < .05$
[73] lack of inversion / lack of auxiliaries in <i>wh</i> -questions	[74] lack of inversion in main clause <i>yes/no</i> questions	$p < .005$
<i>c. L2 Varieties</i>		
[75] <i>like</i> as a focussing device	[76] <i>like</i> as a quotative particle	$p < .05$
[4] regularized reflexives-paradigm	[8] generic <i>he/his</i> for all genders	$p < .05$
[63] relative particle <i>as</i>	[61] relative particle <i>what</i> / [66] gapping or zero-relativization in subject position	$p < .05$
<i>d. English-Based Pidgins and Creoles</i>		
[38] levelling of preterite/ppt verb forms: past replacing the participle	[39] levelling of preterite/ppt verb forms: participle replacing the past form	$p < .01$
[48] invariant <i>don't</i> for all persons in the present tense	[44] multiple negation / negative concord	$p = .05$

Exhibited in the database: if a variety has feature 1, it also has feature 2, but not necessarily vice versa. Rightmost column provides results of one-tailed Fisher's Exact tests on corresponding 2×2 tetrachoric tables.

loosening of sequence of tense rule

		attested	not attested
<i>would in if-clauses</i>	attested	NfldE, . . .	—
	not attested	East Anglia, . . .	CollAmE, . . .

Newfoundland English has both [30] and [31], Colloquial American English has neither, and East Anglia has [30] but not [31]. All of these configurations are predicted by the one-way implication. What is not predicted is a variety that has [31] (*would in if-clauses*) but not [30] (loosening of the sequence of tense rule); indeed, such a variety is not attested in our database. A possible interpretation of this particular one-way implication seems to be that loosening of the sequence of tense rule triggers, or is a necessary precondition to, *would in if-clauses*; the phenomenon, in short, appears to be a specific manifestation of the loosening-of-the-sequence-of-tense-rule phenomenon.

In cross-linguistic typology, one-way implications are often interpreted in terms of historical evolution. This interpretation appears to be appropriate for the implications obtaining between *ain't* as generic negator before a main verb and *ain't* as the negated form of *be/have*. We had seen before that in New Englishes there is an equivalence among the three uses of *ain't*. We can now refine this analysis by stating that if L1 vernaculars are added to the picture, we obtain a preference such that we will not find *ain't* as generic negator before a main verb [47] unless we also see the two more restricted uses of *ain't* [45/46]. This may well have a diachronic explanation such that a L1 vernacular that displays *ain't* as a generic negator must have developed the more specific uses of *ain't* first. This line of argument squares nicely with Anderwald (2003: 48), who, in her book-length treatment of nonstandard negation in the British Isles, argues that generic *ain't* (more specifically, *ain't* for *do*) is an extension of *ain't* for *be* and *have*.

The second bundle of implications in Table 2.2a falls into the domain of relativization strategies. It turns out that any given variety will not have the relative particle *as* [63] unless that variety also has (i) the relative particle *what* [61] and (ii) gapping or zero relativization in subject position [66]. This implication, too, is not entirely unexpected given what we know about the typology of relativization strategies in dialects of English: Zero relativization, for instance, is known to be a precondition to having *as* or *what* as relative markers (Herrmann 2003: 91).

This line of analyzing one-way implications also presents an opportunity to test a hypothesis to be found in the literature on vernacular universals predicting that if a variety exhibits the *was-weren't* split [51], it will also have *was/were* generalization [59], but not vice versa (cf. Britain 2002: 19; Chambers 2004: 132). This should be the case because *was/were* generalization is the more basic phenomenon that may—but need not—develop into the *was-weren't* split. Indeed, we do find evidence for this preference in our

survey: Although the one-way implication between [51] and [59] is not perfect and has an exception—CollAusE has [51] but not [59]—the resulting tetrachoric table is statistically significant at $p = .050$.

Table 2.2b reports a number of one-way implications that hold true for English L1 vernaculars only. For instance, a L1 variety needs habitual *do* [23] in its inventory in order for *do* to be attested as a tense and aspect marker [27]. Once again, this is in line with what we know from the dialectological literature (e.g. Kortmann 2004a). In like manner, *ain't* as the negated form of *be* [45] always implies multiple negation [44]. *Be* as perfect auxiliary [26] is a necessary (but not sufficient!) condition for habitual *be* [22], and lack of inversion in main clause *yes/no* questions [74] can trigger lack of inversion/lack of auxiliaries in *wh*-questions [73], but not the other way around. This is an interesting preference because we had seen in Table 2.1c that this specific pairing is actually an equivalence in English-based pidgins and creoles. In English L1 vernaculars, by contrast, the pairing is a one-way street. Turning to some selected one-way implications in L2 varieties of English (Table 2.2c), we note, for one thing, that *like* as a quotative particle [76] implies *like* as a focusing device [75]. Furthermore, the possibility of regularized reflexives paradigms [4] appears to necessitate generic *helhis* for all genders [8]. Last, Table 2.2d presents a couple of one-way implications that are significant in English-based pidgins and creoles: For one thing, for a pidgin or creole to exhibit past forms of irregular verbs replacing participle forms [38], it must also exhibit participle forms replacing the past form [39]. Second, not unlike L1 varieties requiring multiple negation to have *ain't* for *be*, English-based pidgins and creoles exhibit invariant *don't* [48] only if they also exhibit multiple negation [44].

5. PARAMETERS OF VARIATION IN WORLD ENGLISHES: A BIRD'S EYE PERSPECTIVE

In this section, we adopt a bird's eye perspective to explore what generalizations can be made about underlying dimensions of morphosyntactic variance in World Englishes, drawing on *principal component analysis* (cf. Kim and Mueller 1978) of our database. Principal component analysis is a statistical technique that reduces a number of independent variables to a smaller number of hypothetical constructs, or dimensions, known as 'components' or 'factors', which can be assigned meaningful interpretations by the analyst. What counts in principal component analysis, then, is the arrangement of data along—ideally—meaningful dimensions (or components). In linguistics, principal component analysis and similar techniques have been utilized, for instance, in studies of register variation (cf. Biber 1988, who uses *factor analysis*, a close relative to principal component analysis). Our aim here is to probe where principal component analysis can take us in terms of the *Handbook's* survey: which bundles of morphosyntactic features account, in

concert, for the most variance between varieties of English? Treating varieties as variables and features as cases, we used principal component analysis (with Varimax rotation and Kaiser normalization) to extract two components. The first component accounts for 23.2% of the observable variance in the database, whereas component 2 explains 15.2% of the variance. This means that component 1 is statistically more robust than component 2 and that the two components cumulatively capture c. 38% of the variance⁵ in our survey.

Figure 2.4 visualizes the factor scores for both of our principal components in a two-dimensional coordinate plane. As for the linguistic interpretation of the two components, we suggest that component 1 displays increased levels of ‘morphosyntactic complexity’. Although we are fully aware that ‘complexity’ is a notoriously controversial and ill-defined notion, we tentatively define *morphosyntactic complexity* as follows: A given variety *X* is morphosyntactically more complex than a given variety *X'* if variety *X* exhibits (i) fewer features that arguably simplify syntactic rules, (ii) a smaller amount of features that can be said to aid processing, and (iii) more features that are indicative of ‘distinctions beyond communicative necessity’ (notice that this latter part of our definition heavily draws on McWhorter 2001: 125).

Further, we believe component 2 to indicate a given variety’s degree of ‘analyticity’, a notion which we operationally define as bringing about a greater number of features that are rather autonomous—that is, invariable and periphrastic—in nature, much along the lines of Vincent’s concise definition:

Construction *C* is relatively more analytic than another construction *C'* having approximately the same grammatical content as *C* to the extent that the constituent elements of *C* show greater morphosyntactic and phonological autonomy than do those in *C'*.

(Vincent 1997: 99)

What is the evidence for this specific interpretation of the axes in Figure 2.4? It is our supposition that those features characteristic of varieties towards the right pole of the *x*-axis (i.e. component 1s) are, more often than not, indicative of increased levels of morphosyntactic complexity, and that many of those features that are characteristic of varieties scoring high on the *y*-axis (i.e. component 2) are rather analytic in nature. Thus, as one moves upwards on the vertical axes, one increasingly finds varieties attesting analytical and invariant features such as zero past-tense forms [40], invariant present-tense forms [53], and *never* as a preverbal past-tense negator. As one moves rightwards on the horizontal axis, one increasingly encounters varieties that attest morphosyntactically more complex features such as double comparatives and superlatives [19], inverted word order in

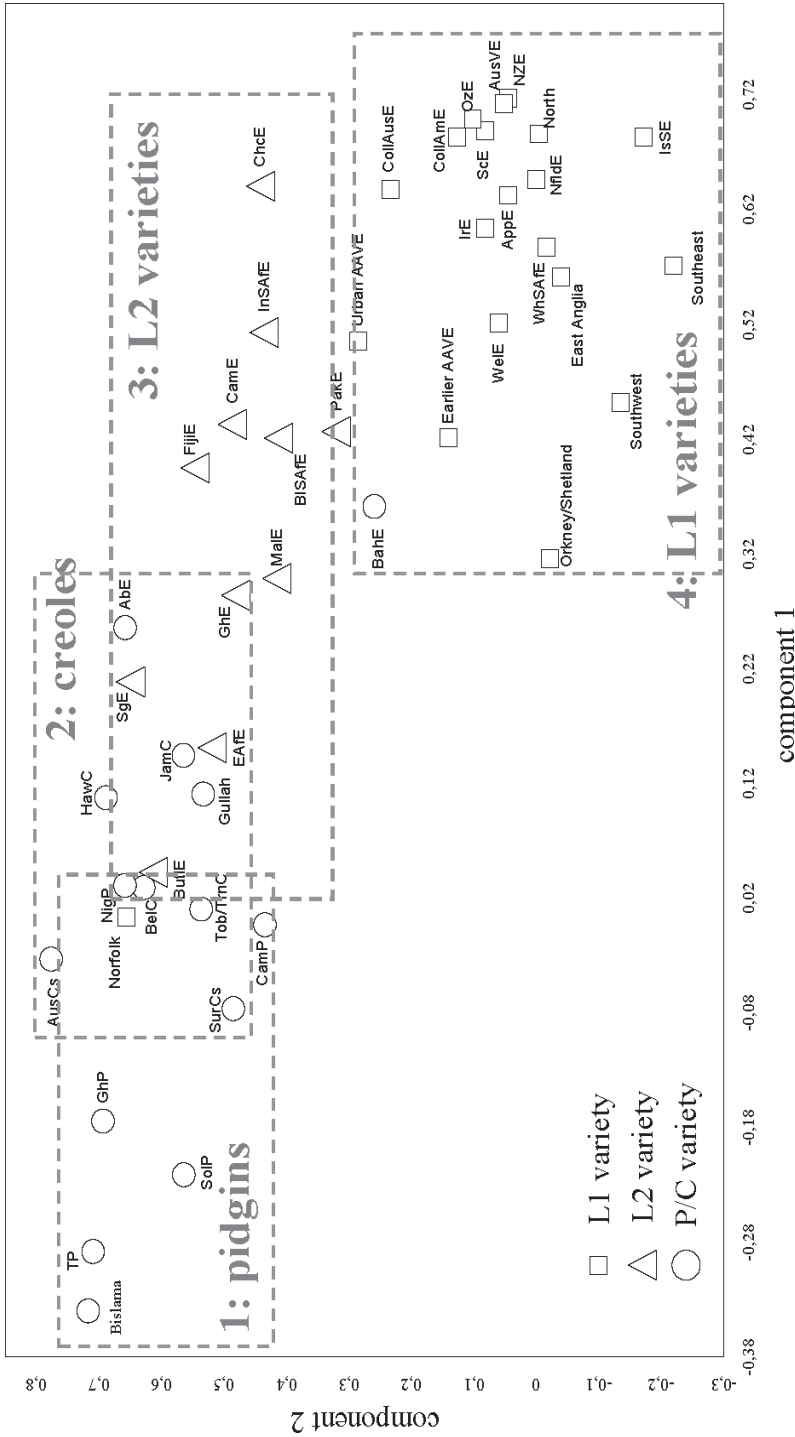


Figure 2.4 Visualization of principal components of variance in the 76 x 46 database. Squares represent L1 varieties, triangles represent L2 varieties, circles represent English-based pidgins and creoles, and dotted boxes indicate statistically significant group membership.

indirect questions [69], and multiple negation [44] (for a detailed discussion, qualitative and statistical, see Szmrecsanyi and Kortmann, forthcoming). What is important here is that visual inspection of Figure 2.4 reveals four more or less distinct groupings, which are indicated by dotted boxes in Figure 2.4: English-based pidgins (box 1), English-based creoles (box 2), L2 varieties of English (box 3), and L1 English vernaculars (box 4). This pattern is neat and faces two outliers only: Norfolk, classified as an L1 variety in our survey, is located in the pidgin/creole box; and Bahamian English (BahE), classified as a pidgin/creole variety, ends up in the L1 box (although only marginally so). Consider, now, that the classification of Norfolk and BahE is actually not as clear-cut as the survey's classification (which was the product of lively discussions among the *Handbook's* editors) might suggest. In his survey article on Norfolk, Mühlhäusler (2004: 789) notes that this variety "shares the characteristic of many creoles, koinés and mixed languages of not having a great deal of inflectional and derivational morphology", whereas Burridge (2004: 1116) classifies Norfolk as a contact variety. Reaser and Torbert (2004: 392–3), in contrast, devote a whole section of their survey article on BahE to the question as to why the linguistic status of the varieties that the label refers to is a problematic one. At any rate, it is well known that the settlement history of the Bahamas is rather unique.

These two outliers notwithstanding, with regard to morphosyntactic complexity (component 1), the hierarchy in (1) emerges:

- (1) English-based pidgins < English-based creoles < L2 varieties of English < English L1 vernaculars

Is this hierarchy meaningful from a theoretical perspective? We would like to argue that it is. As for (morpho-)syntactic complexity, we would like to draw attention to an ongoing debate in the literature about what (if anything) it is that sets apart creoles, as a synchronic class, from other languages. According to McWhorter,

if all of the world's languages could be ranked on a scale of complexity, there would be a delineable subset beginning at the 'simplicity' end and continuing towards the 'complexity' one all of which were creoles.

(McWhorter 2001: 162)

McWhorter argues that this is because

[c]reole languages are unique in having emerged under conditions which occasioned the especial circumstance of stripping away virtually all of a language's complexity . . . such that the complexity emerging in

a creole is arising essentially from ground zero, rather than alongside the results of tens of thousands of years of other accretions. As such, creoles tend strongly to encompass a lesser degree of complexity than any older grammar.

(McWhorter 2001: 155)

Although these statements are not uncontroversial, especially not among the creolist community, our modest empirical inquiry can be seen to suggest that McWhorter's view may not be entirely implausible (at least as far as varieties of English and English-based creoles are concerned). Conflating the English-based pidgins in our survey (box 1) with the English-based creoles (box 2), one finds that in Figure 2.4 all of the leftmost varieties are English-based pidgins and creoles, with Bislama, Tok Pisin, and so on constituting the most extreme cases. Moving rightwards, one then finds overlap areas among pidgins, creoles, and L2 varieties of English: Nigerian Pidgin English, for example, scores as high on the complexity (i.e. horizontal) dimension as Belizean Creole and Butler English; According to Hosali (2004: 1032), an L2 variety, it is hard to say whether ButE should be considered a pidgin or an early fossilized interlanguage). Where boxes 2 and 3 overlap, we find creoles, such as Aboriginal English, that are as complex as some L2 varieties, such as Ghanaian English. Last, one finds a number of L2 varieties—Chicano English, Indian South African English, and so on—that are as complex as the average L1 vernacular, although there also appear to be L1 vernaculars (e.g. Orkney and Shetland English) that appear to be less complex morphosyntactically than some L2 varieties.

What can be said about the other dimension in Figure 2.4? Analyticity (component 2) yields the following hierarchy:

- (2) English-based pidgins / English-based creoles / L2 varieties of English > English L1 vernaculars

This is another way of saying that, in terms of the vertical axis (and, thus, in terms of analyticity), there is not much variance among English-based pidgins, English-based creoles, and L2 varieties. Crucially, however, these three groups score (the usual exceptions, e.g. BahE, apply) higher on the analyticity dimension than any L1 vernacular in the sample. It is well known in this context that creoles in general tend to be highly analytic if not isolating languages (cf. Hagège 1985; McWhorter 1998; Mufwene 1990; Schuchardt 1979). Similarly, the adult L2 acquisition process is known to be hostile toward inflectional morphology (cf. Klein and Perdue 1997), a fact which might ultimately be responsible for the analytic/isolating nature of creole languages (cf. DeGraff 1997). At any rate, given the literature, it is reasonable that among the varieties sampled in our database, L1 vernaculars should be the least analytic ones.

6. SUMMARY AND CONCLUSIONS

Although we are fully aware that the quantitative approach adopted in this study can never replace, but only complement, an in-depth qualitative dialectological inquiry, we hope that this study has succeeded in highlighting the many ways in which methods of typological analysis can be brought to bear on a database mapping language-internal variation. For one thing, we sought to demonstrate that an analytical distinction between different types of versals does more than just proliferate terminology—rather, these are distinctions that, we believe, are indispensable for achieving descriptive and explanatory accuracy. Recall here that in terms of unrestricted versals in particular, we have shown that not everything that is widespread in one particular world region or in one particular variety type is necessarily widespread in other varieties of English, and not everything that is recurrent in varieties of English must necessarily be universal or even pervasive cross-linguistically. In this spirit, we have argued that ‘vernacular universal’, for instance, is a rather strong term that should be used with caution. We also aimed to demonstrate that instructive generalizations can follow not only from considering absolute frequencies, but also from co-occurrence patterns. Carrying the distinction between angloverals and varioversals over to the study of implicational tendencies, we have suggested that different bundles of features, or biconditional implications, characterize different groups of varieties, and that the set of observable one-way implications appears to be conditioned on the variety type under analysis. Such differences may well reflect different pathways of historical development and grammaticalization, an issue which—we believe—is certainly worth paying attention to in future study. Finally, drawing on a principal component analysis of our survey, we argued that varieties of English can be considered to vary along two fundamental dimensions: morphosyntactic complexity and analyticity. These dimensions yield hierarchies which arrange varieties of English according to variety type, rather than according to other factors. Crucially, these hierarchies can, we believe, be considered angloverals, or ‘angloveral continua’, in their own right. Should similar patterns emerge in the study of vernacular varieties of other languages, we are likely to be dealing with cross-linguistically robust *vernacular continua*.

APPENDIX: THE FEATURE CATALOGUE

Note: For a version of the feature catalogue annotated with linguistic examples, see Kortmann and Szmrecsanyi (2004: 1146–8).

Pronouns, Pronoun Exchange, and Pronominal Gender

them instead of demonstrative *those*

me instead of possessive *my*

special forms or phrases for the second person plural pronoun
regularized reflexives paradigm
object pronoun forms serving as base for reflexives
lack of number distinction in reflexives
she/her used for inanimate referents
generic *he/his* for all genders
myself/meself in a nonreflexive function
me instead of *I* in coordinate subjects
nonstandard use of *us*
noncoordinated subject pronoun forms in object function
noncoordinated object pronoun forms in subject function

Noun Phrase

absence of plural marking after measure nouns
group plurals
group genitives
irregular use of articles
postnominal *for*-phrases to express possession
double comparatives and superlatives
regularized comparison strategies

Verb Phrase: Tense and Aspect

wider range of uses of the progressive
habitual *be*
habitual *do*
nonstandard habitual markers other than *do*
levelling of difference between present perfect and simple past
be as perfect auxiliary
do as a tense and aspect marker
completive/perfect *done*
past tense/anterior marker *been*
loosening of sequence of tense rule
would in if-clauses
was sat/stood with progressive meaning
after-perfect

Verb Phrase: Modal Verbs

double modals
epistemic *mustn't*

Verb Phrase: Verb Morphology

levelling of preterite and past participle verb forms: regularization of
irregular verb paradigms

levelling of preterite and past participle verb forms: unmarked forms
levelling of preterite and past participle verb forms: past form replacing the participle
levelling of preterite and past participle verb forms: participle replacing the past form
zero past-tense forms of regular verbs
a-prefixing on *ing*-forms

Adverbs

adverbs (other than degree modifiers) have same form as adjectives
degree modifier adverbs lack *-ly*

Negation

multiple negation/negative concord
ain't as the negated form of *be*
ain't as the negated form of *have*
ain't as generic negator before a main verb
invariant *don't* for all persons in the present tense
never as preverbal past-tense negator
no as preverbal negator
was-weren't split
invariant nonconcord tags

Agreement

invariant present-tense forms due to zero marking for the third person singular
invariant present-tense forms due to generalization of third person *-s* to all persons
existential/presentational *there's*, *there is*, and *there was* with plural subjects
variant forms of dummy subjects in existential clauses
deletion of *be*
deletion of auxiliary *have*
was/were generalization
Northern Subject Rule

Relativization

relative particle *what*
relative particle *that* or *what* in nonrestrictive contexts
relative particle *as*
relative particle *at*
use of analytic *that his/that's*, *what his/what's*, *at's*, *as'* instead of *whose*

52 Benedikt Szmrecsanyi and Bernd Kortmann

gapping or zero relativization in subject position
resumptive/shadow pronouns

Complementation

say-based complementizers
inverted word order in indirect questions
unsplit *for to* in infinitival purpose clauses
as what/than what in comparative clauses
serial verbs

Discourse Organization and Word Order

lack of inversion/lack of auxiliaries in *wh*-questions
lack of inversion in main clause *yes/no* questions
like as a focussing device
like as a quotative particle

NOTES

1. Some readers might contend that copula absence before NPs is quite rare even in varieties of English (cf. Labov 1969: 731 on constraints on copula deletion in AAVE). In this connection, it should be noted that in varieties of English such as AAVE, copula deletion before NPs may be rare but possible.
2. If it actually has such a tie, ‘vernacular typoversal’ might be a better term for a candidate feature.
3. Having said that, we shall see in section 4.2 that in L1 vernaculars, a one-way implication holds between habitual *do* and *do* as a tense and aspect marker.
4. The specific statistical measure used is *Fisher’s Exact test* (cf. Cysouw 2003: 91–2).
5. Statistically speaking, this value is fairly good—by way of comparison, the first two factors in Biber (1988: 83) account for 26.8% and 8.1%, respectively, of the shared variance.

REFERENCES

- Anderwald, L. 2003. *Negation in Non-Standard British English: Gaps, Regularizations and Asymmetries* (Routledge Studies in Germanic Linguistics). London/New York: Routledge.
- Barbiers, S., L. Cornips, and S. van der Kleij, eds. 2002. *Syntactic Microvariation*. Amsterdam: SAND.
- Biber, D. 1988. *Variation Across Speech and Writing*. Cambridge: Cambridge University Press.
- Britain, D. 2002. Diffusion, leveling, simplification and reallocation in the past tense of BE in the British Fens. In *Investigating Change and Variation Through Dialect Contact*, Special Issue of *Journal of Sociolinguistics* 6, edited by L. Milroy, 16–43.
- Burridge, K. 2004. Synopsis: Morphological and syntactic variation in the Pacific and Australasia. In Kortmann et al., eds. 2004. Vol 1. 116–1131.

- Chambers, J. K. 2003. *Sociolinguistic Theory: Linguistic Variation and Its Social Implications*. Oxford Malden: Blackwell.
- . 2004. Dynamic typology and vernacular universals. In *Dialectology Meets Typology*, edited by B. Kortmann, 127–45. Berlin and New York: Mouton de Gruyter.
- Cysouw, M. 2003. Against implicational universals. *Linguistic Typology* 7: 89–101.
- DeGraff, M. 1997. Verb syntax in, and beyond, creolization. In *The New Comparative Syntax*, edited by L. Haegeman, 64–94. London: Longman.
- Greenberg, J. H. 1963. *The Languages of Africa*. Bloomington: Indiana University.
- . 1966. *Language Universals, With Special Reference to Feature Hierarchies*. The Hague: Mouton.
- Hagège, C. 1985. *L'homme de paroles*. Paris: Fayard.
- Haspelmath, M., M. Dryer, D. Gil, and B. Comrie, eds. 2005. *The World Atlas of Language Structures*. Oxford: Oxford University Press.
- Herrmann, T. 2003. *Relative clauses in dialects of English: a typological approach*. PhD thesis, University of Freiburg. Available from <http://www.freidok.uni-freiburg.de/volltexte/830/> (accessed 14 November 2006).
- Hosali, P. 2004. Butler English: Morphology and syntax. In Kortmann et al., eds. 2004. Vol. 2. 1031–44.
- Kim, J. O., and C. W. Mueller. 1978. *Introduction to Factor Analysis: What It Is and How To Do It* (Quantitative Applications in the Social Sciences). Newbury Park/London/New Delhi: Sage Publications.
- Klein, W., and C. Perdue. 1997. The basic variety (or: Couldn't natural languages be much simpler?). *Second Language Research* 13: 301–47.
- Kortmann, B. 2004a. Do as a tense and aspect marker in varieties of English. In Kortmann, ed. 2004b. 245–75.
- , ed. 2004b. *Dialectology Meets Typology: Dialect Grammar from a Cross-Linguistic Perspective*. Berlin and New York: Mouton de Gruyter.
- , E. Schneider, K. Burridge, R. Mesthrie, and C. Upton, eds. 2004. *A Handbook of Varieties of English*, 2 vols. Berlin and New York: Mouton de Gruyter.
- and B. Szmrecsanyi. 2004. Global synopsis: Morphological and syntactic variation in English. In Kortmann et al., eds. 2004. Vol. 2. 1142–202.
- Labov, W. 1969. Contraction, deletion and inherent variability of the English copula. *Language* 45: 715–62.
- McWhorter, J. 1998. Identifying the creole prototype: Vindicating a typological class. *Language* 74: 788–818.
- . 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* 6: 125–66.
- Mühlhäusler, P. 2004. Norfolk Island-Pitcairn English (Pitkern Norfolk): Morphology and syntax. In Kortmann et al., eds. 2004. Vol. 2. 789–801.
- Mufwene, S. 1990. Transfer and the substrate hypothesis in creolistics. *Studies in Second Language Acquisition* 12: 1–23.
- Reaser, J., and B. Torbert. 2004. Bahamian English: Morphology and syntax. In Kortmann et al., eds. 2004. Vol. 2. 391–406.
- Schuchardt, H. 1979. *The Ethnography of Variation: Selected Writings on Pidgins and Creoles*. Ann Arbor, MI: Karoma.
- Stassen, L. 2005. Zero copula for predicate nominals. In Haspelmath et al., eds. 2005. 486–9.
- Szmrecsanyi, B., and B. Kortmann. Forthcoming. The morphosyntax of varieties of English worldwide: A quantitative perspective.
- Vincent, N. 1997. Synthetic and analytic structures. In *The Dialects of Italy*, edited by M. Maiden, 99–105. London: Routledge.