Benedikt Szmrecsanyi

Recontextualizing language complexity

Abstract: There is currently much interest in language complexity, but the research community relies excessively on measures of system complexity, giving short shrift to usage. Against this backdrop I sketch three ways to measure usage complexity in actual text and speech: typological profiles (which explore the extent to which languages and lects use synthetic or analytic grammatical marking), Kolmogorov complexity (which is about the predictability of new text given knowledge about old text), and variational complexity (which is concerned with the complexities of choosing between linguistic variants). I argue in conclusion that the study of language and (dia)lect complexity in context would mesh well with the spirit of the cognitive sociolinguistics paradigm.

1 Introduction

One of the defining characteristics of the productive research program launched and directed by Dirk Geeraerts is its emphasis on the usefulness and, in fact, indispensability of usage data (see Geeraerts 1985: 29 for an early reference). In precisely this spirit, this contribution seeks to demonstrate how language and (dia)lect complexity may be operationalized as a usage-based concept.

My point of departure is that research on language complexity is currently booming. There is an emerging consensus that human languages, and dialects of the same language, may differ with regard to their complexity. The trouble is that the plethora of measures used in the literature to measure complexity are concerned with Saussurean *langue, or* "system complexity", in the parlance of Pallotti (2015). For example, analysts have counted the number of contrastive elements in a system (Nichols 2013), the number of rules in a grammar (McWhorter 2001), or have been interested in whether or not a language has

Benedikt Szmrecsanyi: University of Leuven

grammatical gender (Trudgill 1999). These are all fine complexity indicators, but they do selectively restrict attention to LANGUAGE STRUCTURE and KNOWLEDGE[1].

Against this backdrop, this contribution presents ways to consider LANGUAGE USAGE in theoretically oriented complexity research. I will be specifically concerned with three complexity measures: (1) typological profiles – the extent to which languages use synthetic or analytic grammatical marking, (2) Kolmogorov complexity – the extent to which new text is predictable from old text, and (3) variational complexity – the extent to which choosing between linguistic variants is subject to restrictions.

This paper is structured as follows. In section 2, I briefly summarize the history of thought on language complexity. Section 3 presents the three usage-based complexity measures. Section 4 offers some concluding remarks.

2 A brief history of thought on language complexity

For most of the twentieth century, many linguists agreed that all languages are equally complex, an article of faith that has been dubbed the *linguistic equi-complexity dogma* (Kusters 2003: 5). This consensus has been eroding in recent years. One of the primers was a lead article in the journal *Linguistic Typology* in which John McWhorter suggested, a bit provocatively, that creole languages tend to have simpler grammars than older languages, "by virtue of the fact that they were born as pidgins, and thus stripped of almost all features unnecessary to communication" (McWhorter 2001: 125).

Most recent work on language complexity takes a functional-typological (e.g. Miestamo, Sinnemäki, and Karlsson 2008), contact linguistic (e.g. Kortmann and Szmrecsanyi 2012), and/or sociolinguistic (e.g. Trudgill 2011) perspective. A crucial theme in this literature concerns the question of how to

1 I hasten to add that by contrast to theoretical linguistics, applied linguists and SLA researchers have been in the business of measuring complexity for a long time. Customary complexitymeasures in this field include the length of syntactic units, density of subordination, and the frequency of occurrence of "complex forms". While there measures are certainly nicely amenable to operationalization in usage data, they do suffer from "concept reductionism" (Ortega 2012: 128) and will hence not be explored further in this contribution.

best measure complexity. Many analysts distinguish between measures of ABSOLUTE COMPLEXITY (theory-driven) and measures of RELATIVE COMPLEXITY (about "difficulty"; see Miestamo 2009 for discussion). Absolute complexity measures include ABSOLUTE-QUANTITATIVE COMPLEXITY (e.g. the length of the minimal description of a linguistic system; Dahl 2004), REDUNDANCY-INDUCED COMPLEXITY (a.k.a. "baroque accretion"; see McWhorter 2001: 126), or IRREGULARITY-INDUCED COMPLEXITY (see, e.g., Nichols 2013). As for relative complexity, analysts have primarily been interested in L2 ACQUISITION COMPLEXITY, defined by Kusters (2003: 6) as "the amount of effort an outsider has to make to become acquainted with the language in question" (see also Trudgill 2001: 371). Despite taking the language user as their reference point, I stress that relative complexity measures are not necessarily usage-based; they may very well – and indeed often are – interested in structural aspects that may make a language user-friendly.

The second major theme in the literature is the extent to which complexity variation is a function of social factors. Research in this vein typically assumes that languages are complex adaptive systems (Beckner et al. 2009) whose complexity profiles adapt to the communicative, cultural, and cognitive needs of their speakers (Bentz and Winter 2013: 19). Hence language structure and its complexity are a function of social and/or sociohistorical factors. For example, it has been argued that a history of language contact and concomitant simplificatory adult SLA triggers simplification (Trudgill 2001: 372; Lupyan and Dale 2010). Conversely, complexification is thought to occur in contact scenarios that involve childhood bilingualism (Trudgill 2011: 42) and in the absence of contact (Nichols 2013; Wray and Grace 2007). History aside, it is well known that complexity can be a function of the speech situation: registers differ in terms of the extent and type(s) of complexity that they exhibit (Biber, Gray, and Poonpon 2011).

3 How to measure complexity in usage data

The upshot is, then, that while linguists have become quite good at determining and interpreting system complexity on the level of *langue*, complexity of *parole* (i.e. usage complexity, or text complexity) has received comparatively short shrift. Part of the reason why this is the case is that system complexity can be determined by consulting reference grammars (the primary data source in typological research), while usage complexity can only be measured in corpora sampling actual text and speech (which are often not available, particularly not in the case of less well documented languages). But when corpora sampling

language usage are available, we can and should measure usage complexity. I what follows I will outline three ways to do just that.

3.1 Typological profiling

ANALYTICITY (the extent to which languages use free markers and word order to signal grammatical relationships) and SYNTHETICITY (the extent to which languages rely on bound markers, such as inflections) are two time-honored (see, e.g., Schlegel 1818) holistic typology notions. What is important here is that analyticity and syntheticity are often interpreted in terms of complexity: analytic marking is usually thought of as increasing explicitness and transparency (Humboldt 1836: 284–285), whereas synthetic marking is seen as difficult thanks to the allomorphies it creates (Braunmüller 1990: 627). It is also a known interlanguage universal that learners avoid inflectional marking and prefer analyticity (e.g. Klein and Perdue 1997). In short, analyticity counts as simple, syntheticity as complex. Crucially, Greenberg (1960) demonstrated that seemingly abstract typological notions such as analyticity and syntheticity are actually amenable to precise measurements on the basis of actual texts. Drawing inspiration from Greenberg's method, I will show in this section how we can calculate indices to profile the way in which grammatical information is coded in usage data.

Given a part-of-speech annotated corpus, we can use the annotation to group word tokens in corpus texts into three broad categories: (1) analytic word tokens, i.e. function words that are members of synchronically closed word classes (e.g. conjunctions, determiners, pronouns, prepositions, modal and auxiliary verbs, negators); (2) synthetic word tokens, which carry bound grammatical markers (e.g. inflectionally marked verbs and nouns); and (3) simultaneously analytic *and* synthetic word tokens (e.g. inflected auxiliary verbs). Once this categorization is in place, we can calculate two Greenberg-inspired indices: the Analyticity Index, which is calculated as the ratio of the number of free grammatical markers (i.e. function words) in a text to the total number of words in the text, normalized to a sample size of 1,000 words of running text; and the Syntheticity Index, which is calculated as the ratio of the number of words in a text that bear a bound grammatical marker to the total number of words in the sample text, normalized to a sample size of 1,000 words of running text. (For a more detailed description of the method, see Szmrecsanyi 2009.)

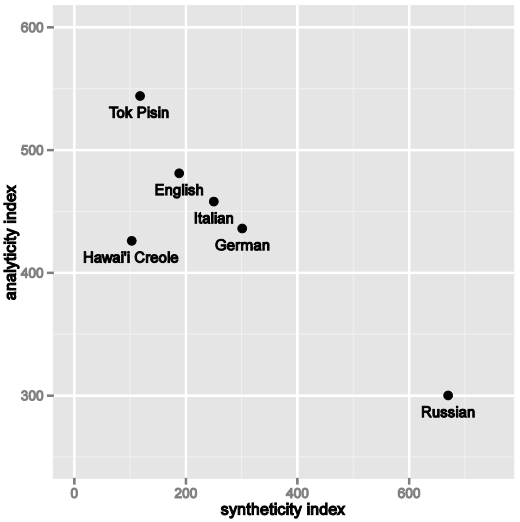

Fig. 1: Analyticity Index scores (*y*-axis) against Syntheticity Index scores (*x*-axis) in European languages and two English-based creole languages (adapted from Siegel, Szmrecsanyi, and Kortmann 2014, Figure 1).

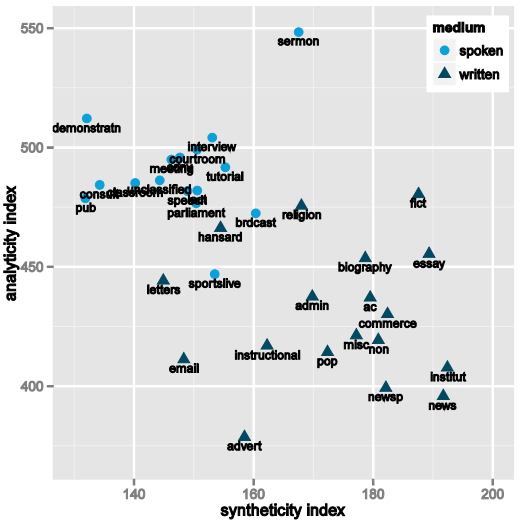

Fig. 2: Analyticity Index scores (*y*-axis) against Syntheticity Index scores (*x*-axis) of text types sampled in the British National Corpus (adapted from Szmrecsanyi 2009, Figure 2).

In what follows I showcase the descriptive and interpretative benefits of this approach. Using the method, Siegel, Szmrecsanyi and Kortmann (2014) create typological profiles of a number of European languages (English, Italian, German, and Russian) as well as two English-based creole languages (Tok Pisin and Hawai'i Creole), tapping into a corpus of written texts. Figure 1 thus locates the data points in a two-dimensional analyticity-syntheticity plane. It turns out that Russian is the most synthetic and least analytic language in the sample, while Tok Pisin is the most analytic and least synthetic language (Hawai'i Creole is also fairly non-synthetic, but less analytic than Tok Pisin). Equating analyticity with simplicity and syntheticity with complexity – as is customary in the literature – this is another way of saying that Tok Pisin, Hawai'i Creole and English are the least complex languages in the sample while Russian is the most complex language. Figure 2 applies the method to the various spoken and written text types sampled in the British National Corpus (BNC). Here, we find a very clear split between spoken text types and written text types: spoken text types are more analytic and less synthetic than written text types. In terms of complexity, we would thus conclude that spoken text types are less complex than written text types – which makes sense, given that spoken language is subject to all kinds of online processing constraints in a way that written language is not.

3.2 Kolmogorov complexity

Typological profiles are informative, but they do restrict attention to aprioristically defined "interesting" dimensions of grammatical variability. The usage-based complexity measure that we will be discussing now, KOLMOGOROV COMPLEXITY, does no such thing. As a measure that brings in information theory (Shannon 1948), Kolmogorov complexity is unsupervised, holistic, and radically text-based: Kolmogorov complexity defines the complexity of a string or text as the length of the shortest possible description of that string or text.

To my knowledge Juola (1998, 2008) was the first to utilize Kolmogorov complexity in the realm of language complexity research. His idea was that text samples that can be compressed efficiently are linguistically simple, while texts that cannot be compressed efficiently are complex. This is another way of saying that according to the measure, texts are linguistically simple or complex to the extent that they can or cannot be predicted from previously seen texts. It is clear that Kolmogorov complexity is entirely agnostic about form-meaning relationships and such things; what is measured is text-based linguistic surface complexity/redundancy (see Ehret 2014; Ehret in preparation for extended dis-

cussion). But because it analyzes texts (and not, e.g., grammar books), Kolmogorov complexity is a usage-based measure.

A nice property of the Kolmogorov complexity is that it can be conveniently approximated using off-the-shelf file compression programs. These use adaptive entropy estimation, which approximates Kolmogorov complexity (Juola 1998). File compression programs compress text strings by describing new strings on the basis of previously seen and memorized (sub-)strings so that the amount of information and redundancy in a given string can be measured (Juola 2008: 93). We may measure the overall Kolmogorov complexity of a text using the following procedure: (1) feed corpus texts into a compression program such as *gzip* (the results to be reported below were obtained using *gzip* version 1.2.4), (2) note down file sizes before and after compression, (3) regress out the trivial correlation between the two measures, (4) interpret the regression residuals (in bytes) as adjusted complexity scores: bigger adjusted complexity scores indicate more Kolmogorov complexity[2].

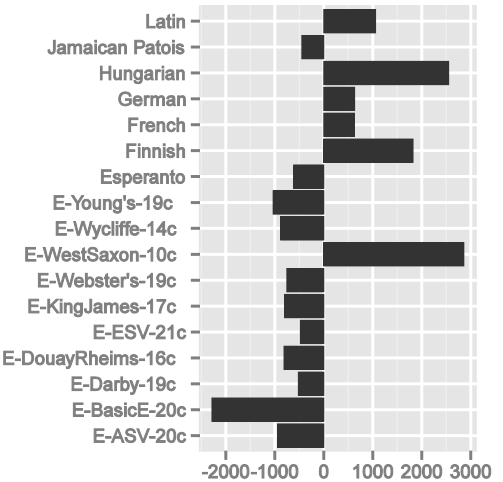

Fig. 3: Kolmogorov complexity of Bible texts: adjusted overall complexity scores by Bible translation. Negative adjusted complexity scores indicate below-average complexity; positive adjusted complexity scores indicate above-average complexity (adapted from Ehret and Szmrecsanyi in press, Figure 1).

2 In addition, the method may be combined with distortion techniques to address complexity at the morphological and syntactic tier (see Juola 2008; Ehret in preparation for details).

To demonstrate that the method works, Ehret and Szmrecsanyi (in press) measure Kolmogorov complexity in a parallel text database containing translations of the Gospel of Mark into a number of languages (Esperanto, Finnish, French, German, Hungarian, Jamaican Patois, and Classical Latin), including (mostly historical) translations into English (from a West Saxon translation over the King James Bible to the English Standard Version, published in 2001). Figure 3 shows that the Kolmogorov approach ranks the complexity of the Bible texts in a way that seems to be compatible with what we think we know about the languages covered in the sample. The three most complex translations are the West Saxon, Hungarian, and Finnish texts; Jamaican Patois and Esperanto are rather non-complex, and so are most translations into English (except for the West Saxon translation mentioned above). The least complex data point in the sample is the Basic English translation of the Bible. Basic English is a simplified variety of English designed by Charles Kay Ogden as, among other things, an aid to facilitate teaching of English as a foreign language (Ogden 1934). Figure 3 seems to suggest that Ogden did a fairly good job.

3.3 Variational complexity

In this section, I would like to offer some thoughts on how the analysis of linguistic variation can be made relevant to research on language complexity (and vice versa). Variation analysts are interested in the factors that constrain choices between "alternate ways of saying 'the same' thing" (Labov 1972: 188). Recent work on grammatical variation specifically has shown some interest in how variation patterns can be interpreted as being more or less complex.

I illustrate the state-of-the-art in variation analysts by summarizing Bresnan et al. (2007), who explore the alternation between the ditransitive dative pattern, as in (1a), and the prepositional dative variant, as in (1b).

(1) a. *The linguist sent [the President]*$_{\text{recipient}}$ *[a letter]*$_{\text{theme}}$
b. *The linguist sent [a letter]*$_{\text{theme}}$ *[to the President]*$_{\text{recipient}}$

Bresnan et al. extracted all variable dative occurrences from the Switchboard corpus, which samples spoken American English, and annotated the tokens for a large number of predictor variables. On the basis of this annotation, Bresnan et al. then fit regression models to predict speakers' syntactic choices. The models correctly predicted more than 90% of the actual dative choices and showed that the dative alternation in American English is constrained by at least 10

factors – e.g. animacy of the recipient, definiteness of the theme, length of the recipient and theme, and so on.

Bresnan et al. (2007) were not concerned with language complexity, but their findings may be interpreted against this backdrop. Is ten constraints on the dative alternation the last word, or are there simpler (dia)lects where the dative alternation is only constrained by, say, five constraints? Conversely, are there more complex (dia)lects where the dative alternation is constrained by more than ten factors? Taking as point of departure questions like these, a measure of VARIATIONAL COMPLEXITY would define language complexity as being a function of the quantity of constraints on variation. The rationale is that more constrained variational patterns require, on the one hand, more description; on the other hand, more constrained variational patterns presumably harder to acquire than less constrained patterns.

I illustrate this basic idea with the help of three concrete examples. Shin (2014) utilizes the variationist method to study variation in Spanish 3rd person singular subject pronoun expression (*ella canta* versus ___ *canta*). Looking at speakers in New York City, Shin finds that US-born Latinos have more constrained variation grammars than first-generation Latin American immigrants. Specifically, tense/mood/aspect issues constrain the variation between overt and non-overt subject expression in the speech of the second generation, but not the first. Shin argues that US-born Latinos have more constrained variation grammars: "[t]he loss of a linguistic factor that constrains linguistic choice is a type of simplification, while the emergence of a new factor is a type of complexification" (Shin 2014: 305).

Schneider (submitted) studies future marker choice (*I will sit down quietly* versus *I am going to sit down quietly*). She compares the constraint system of Ghanaian English – an indigenized L2 variety presumably subject to simplification pressures, thanks to adult SLA – and native British English. Using annotated corpus-derived datasets, Schneider fits regression models for each of the two varieties she studies. She finds that the minimal adequate model for British English needs five significant constraints (plus a number of interaction terms), while the minimal adequate model for Ghanaian English only has three constraints (clause type, sentence type, and presence of temporal adverbials). Using Shin's (2014) criterion, one would thus conclude that future marker choice in British English is more complex than in Ghanaian English, where it is less constrained.

Szmrecsanyi et al. (under review) study the probabilistic grammar of syntactic variation in four international varieties of English (British English, Canadian English, Indian English, and Singapore English). Among other things,

Szmrecsanyi et al. investigate the particle placement alternation (*Tom looked up the word* versus *Tom looked the word up*). A re-analysis of this particular alternation in terms of variational complexity shows that particle placement variation is subject to fewer constraints in Indian and Singapore English than in the other varieties, according to regression analysis. In addition, it turns out that the constraints also seem to interact less extensively in these varieties than in British and Canadian English. Like Ghanaian English, Indian and Singapore English are non-native varieties of English with a history of adult SLA, and so the observation that they seem to be less complex variationally than native varieties is predicted by theory.

In all, it appears that there are plenty of meaningful variational complexity differentials waiting to be explored. Of course, counting the sheer number of constraints (interpretation: more constraints are more complex) on variation and establishing the extent to which constraints interact (interpretation: interactions induce complexity) are but two ways to assess variational complexity. Other criteria would include determining the relative importance of social constraints (interpretation: social constraints induce complexity), and determining the relative importance of lexically conditioned constrains (interpretation: lexically conditioned constraints induce complexity).

More generally, variational complexity is an innovative concept for a number of reasons, including the following. First, the focus is not on the complexity of linguistic material per se (*substantive complexity*), but on the complexity of linguistic choice-making (*procedural complexity*). Second, the metric is interested not in the presence/absence of "complex" features (*categorical complexity*), but on usage patterns that can be described using the mathematics of uncertainty (*probabilistic complexity*). Third, because variation relates to both usage (*parole*) and knowledge about the system (*langue*) (see Bresnan 2007 for evidence), the approach goes some way towards bridging the gap between system and usage complexity. The drawback is that it is imperative to restrict attention to very well-researched phenomena, as one needs to be sure that whenever one observes complexity differentials, these are not trivially due to the fact that we do not fully understand what constrains the variation at hand.

4 Conclusion

This contribution has surveyed three ways to measure usage complexity: typological profiles (the extent to which languages use synthetic or analytic grammatical marking), Kolmogorov complexity (about the predictability of new text

given knowledge about old text), and variational complexity (about the complexities of choosing between linguistic variants). As I have argued, usage complexity has received short shrift in the previous literature. But my point is not that usage complexity is a "superior" concept vis-à-vis system complexity. Rather, I would like to offer that we need both perspectives to realistically assess language complexity.

Complexity differentials are very interesting from a theoretical point of view. This is because such differentials cannot have biological or communicative reasons: human beings – whatever their native language background – are endowed with the exact same linguistic capacities, and languages – wherever they are spoken – have, on the whole (and especially in "fundamental" registers such as face-to-face conversation), similarly complex or simple functions, such as talking about the weather, and so on. Many analysts thus conclude that complexity variation must have sociolinguistic motives (see, e.g., Trudgill 2011; Wray and Grace 2007). At the same time, language complexity seems advertise itself as a phenomenon to be analyzed in terms of conceptualization and the interplay between usage and knowledge about the system.

This being so, language complexity – and especially so the usage-based sort – is a notion that is, or should be, of interest to students of cognitive sociolinguistics, a framework advanced by Dirk Geeraerts (see, e.g., Geeraerts, Kristiansen, and Peirsman 2010). Cognitive sociolinguistics endeavors to align variational (socio)linguistics and cognitive linguistics, thus "recontextualizing" (Geeraerts 2010) grammar in its social context of use. I believe that there are plenty of exciting avenues for (re-)interpreting language complexity along cognitive sociolinguistics lines.

Acknowledgement

An Odysseus grant by the Research Foundation Flanders (FWO, grant no. G.0C59.13N) is gratefully acknowledged. I thank Jason Grafmiller for reanalyzing the data on the particle placement alternation. The usual disclaimers apply.

References

Beckner, Clay, Richard Blythe, Joan Bybee, Morten H. Christiansen, William Croft, Nick C. Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman & Tom Schoenemann. 2009. Language is a complex adaptive system: Position paper. *Language Learning* 59 (December). 1–26.

Bentz, Christian & Bodo Winter. 2013. Languages with more second language learners tend to lose nominal case. *Language Dynamics and Change* 3. 1–27.

Biber, Douglas, Bethany Gray & Kornwipa Poonpon. 2011. Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly* 45(1). 5–35.

Braunmüller, Kurt. 1990. Komplexe Flexionssysteme – (k)ein Problem für die Natürlichkeitstheorie? *Zeitschrift Für Phonetik, Sprachwissenschaft Und Kommunikationsforschung* 43. 625–35.

Bresnan, Joan. 2007. Is syntactic knowledge probabilistic? Experiments with the English dative alternation. In Sam Featherston & Wolfgang Sternefeld (eds.), *Roots: Linguistics in search of its evidential base* (Studies in Generative Grammar 9), 75–96. Berlin & New York: Mouton de Gruyter.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina & Harald R. Baayen. 2007. Predicting the dative alternation. In Gerlof Boume, Irene Kraemer & Joost Zwarts (eds.), *Cognitive foundations of interpretation*, 69–94. Amsterdam: Royal Netherlands Academy of Science.

Dahl, Östen. 2004. *The growth and maintenance of linguistic complexity*. Amsterdam & Philadelphia: John Benjamins.

Ehret, Katharina. In preparation. Exploring information-theoretic methods to assess linguistic complexity in English. Freiburg: Freiburg University dissertation.

Ehret, Katharina. 2014. Kolmogorov complexity of morphs and constructions in English. *Language Issues in Linguistic Technology* 11(2). 43–71.

Ehret, Katharina & Benedikt Szmrecsanyi. In press. An information-theoretic approach to assess linguistic complexity. In Guido Seiler & Raffaela Baechler (eds.), *Complexity, variation, and isolation*. Berlin: de Gruyter.

Geeraerts, Dirk. 1985. *Paradigm and paradox: Explorations into a paradigmatic theory of meaning and its epistemological background*. Leuven: Leuven University Press.

Geeraerts, Dirk. 2010. Recontextualizing Grammar: Underlying trends in thirty years of cognitive linguistics. In Elzbieta Tabakowska, Michal Choinski & Lukasz Wiraszka (eds.), *Cognitive linguistics in action: From theory to application and back* (Applications of Cognitive Linguistics 14), 71–102. Berlin & New York: De Gruyter Mouton.

Geeraerts, Dirk, Gitte Kristiansen & Yves Peirsman (eds). 2010. *Advances in cognitive sociolinguistics*. Berlin & Boston: Mouton de Gruyter.

Greenberg, Joseph H. 1960. A quantitative approach to the morphological typology of language. *International Journal of American Linguistics* 26(3). 178–94.

Humboldt, Wilhelm von. 1836. *Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwicklung des Menschengeschlechts*. Berlin: Dümmler.

Juola, Patrick. 1998. Measuring linguistic complexity: The morphological tier. *Journal of Quantitative Linguistics* 5(3). 206–13.

Juola, Patrick. 2008. Assessing linguistic complexity. In Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds.), *Language complexity: Typology, contact, change*, 89–108. Amsterdam & Philadelphia: Benjamins.

Klein, Wolfgang & Clive Perdue. 1997. The basic variety (or: Couldn't natural languages be much simpler?). *Second Language Research* 13. 301–47.
Kortmann, Bernd & Benedikt Szmrecsanyi (eds.). 2012. *Linguistic complexity: Second language acquisition, indigenization, contact* (linguae & litterae 13). Berlin: De Gruyter.
Kusters, Wouter. 2003. *Linguistic complexity: The influence of social change on verbal inflection*. Utrecht: LOT.
Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Philadelphia Press.
Lupyan, Gary & Rick Dale. 2010. Language Structure Is Partly Determined by Social Structure. *PLoS ONE* 5(1). e8559.
McWhorter, John. 2001. The World's simplest grammars are creole grammars. *Linguistic Typology* 6. 125–66.
Miestamo, Matti. 2009. Implicational hierarchies and grammatical complexity. In Geoffrey
Sampson, David Gil & Peter Trudgill (eds.), *Language complexity as an evolving variable*, 80–97. Oxford: Oxford University Press.
Miestamo, Matti, Kaius Sinnemäki & Fred Karlsson (eds.). 2008. *Language Complexity: Typology, Contact, Change* (Studies in Language Companion Series 94). Amsterdam & Philadelphia: Benjamins.
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* (linguae & litterae 24), 38–60. Berlin & New York: Walter de Gruyter.
Ogden, Charles K. 1934. *The system of Basic English*. New York: Harcourt.
Ortega, Lourdes. 2012. Interlanguage complexity: A construct in search of theoretical renewal. In Bernd Kortmann & Benedikt Szmrecsanyi (eds.), *Linguistic complexity* (linguae & litterae 13), 127–155. Berlin & Boston: De Gruyter.
Pallotti, Gabriele. 2015. A simple view of linguistic complexity. *Second Language Research* 31(1). 117–134.
Schlegel, August Wilhelm von. 1818. *Observations sur la langue et la littérature provençales*. Paris: Librairie grecque-latine-allemande.
Schneider, Agnes. Submitted. Aspect and modality in Ghanaian English: A corpus-based study of the progressive and the modal WILL. Freiburg: University of Freiburg dissertation.
Shannon, Claude E. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27. 379–423.
Shin, Naomi Lapidus. 2014. Grammatical complexification in Spanish in New York: 3sg pronoun expression and verbal ambiguity. *Language Variation and Change* 26. 1–28.
Siegel, Jeff, Benedikt Szmrecsanyi & Bernd Kortmann. 2014. Measuring analyticity and syntheticity in Creoles. *Journal of Pidgin and Creole Languages* 29(1). 49–85.
Szmrecsanyi, Benedikt. 2009. Typological parameters of intralingual variability: Grammatical analyticity versus syntheticity in varieties of English. *Language Variation and Change* 21(3). 319–53.
Szmrecsanyi, Benedikt, Jason Grafmiller, Benedikt Heller & Melanie Röthlisberger. Under review. Around the world in three alternations: Modeling syntactic variation in global varieties of English. *English World-Wide*.
Trudgill, Peter. 1999. Language contact and the function of linguistic gender. *Poznan Studies in Contemporary Linguistics* 35. 133–152.
Trudgill, Peter. 2001. Contact and simplification: Historical baggage and directionality in linguistic change. *Linguistic Typology* 5 (2-3). 371–74.

Trudgill, Peter. 2011. *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford & New York: Oxford University Press.
Wray, Alison & George W. Grace. 2007. The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua* 117(3). 543–78.