



Kretzschmar, W. A. (2014) Complex systems in aggregated variation analyses. In: Szmrecsanyi, B. and Wälchli, B. (eds.) *Aggregating Dialectology, Typology, and Register Analysis: Linguistic Variation in Text and Speech*. Series: *Language and litterae* (28). De Gruyter: Berlin, pp. 150-173. ISBN 9783110317558.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

<http://eprints.gla.ac.uk/130575/>

Deposited on: 11 November 2016

Enlighten – Research publications by members of the University of Glasgow
<http://eprints.gla.ac.uk>

Workshop on Cross-Linguistic and Language-Internal Variation in Text and Speech, 2011

Complex Systems in Aggregated Variation Analyses

William A. Kretzschmar, Jr., University of Georgia

The American novelist Walker Percy was not a trained linguist, but he had a passion for language and a strong interest in linguistics. In 1977 he commented on what he saw as the state of our field, framed here in the traditional terms of rational vs. empirical approaches:¹

Scientists [of language] tend to be dogmatic about the nature of man. Again they remind me of the Scholastics battling with Galileo. Scholastics spent thousands of man-hours inside their heads trying to prove that Jupiter couldn't have moons and that the earth was at the center of the universe. To suggest otherwise offended their sense of the order of things. Galileo pointed to his telescope: Why don't you take a look? Today we have plenty Scholastics of language. What we need is a Galileo who is willing to take a look at it.

When Percy died in 1990 the movement in linguistics that followed his advice, usage-based linguistics, was just getting started. Paul Hopper presented his seminal paper called "Emergent Grammar" in Berkeley in 1987. Ronald Langacker published a chapter on "A Usage-Based Model" for cognitive linguistics in 1988. At about the same time, the new science of complex systems, also known as complex adaptive systems or dynamical systems, was just getting off the ground. In 1984, the same year that the Sante Fe Institute was founded for the study of complex systems, Lindblom, MacNeilage, and Studdert-Kennedy published a paper on self-organizing processes in phonology. In my talk today, I would like to trace the development of usage-based linguistics and its association with complex systems by conducting a tour for you of some landmark works in the field. The point of this exercise is to show that work by usage-based linguists and study of complex systems in speech, while making great progress, continue to be handicapped by notions carried over from Percy's "Scholastics of language." There is some irony in the fact that many my authors today argue quite sharply against the generative model as they attempt to defend their usage-based model, and yet they still strive to align what they do with

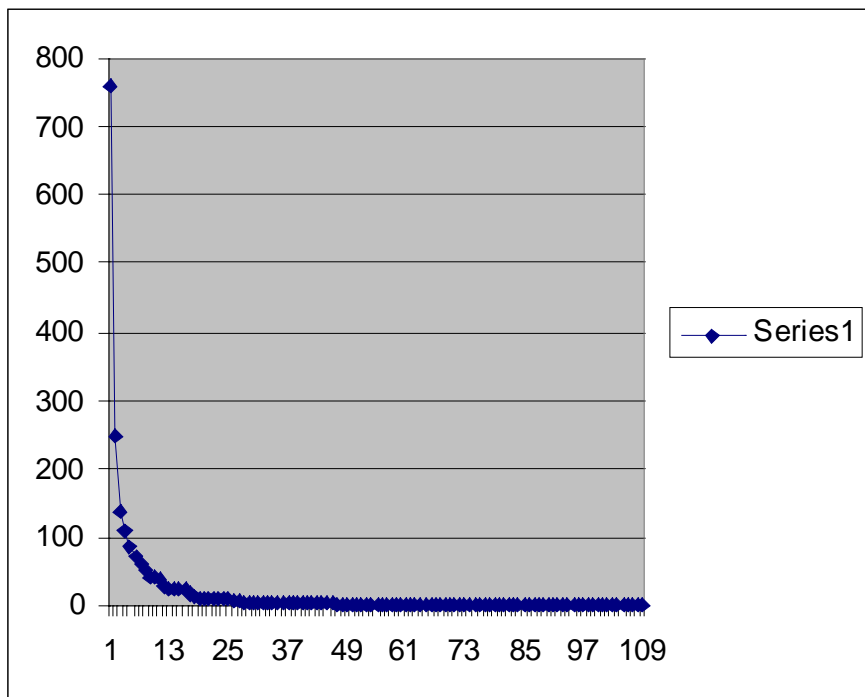
ideas derived from formal linguistics. Usage-based linguistics is just different from generative linguistics, especially when we apply complexity theory in our work. I believe that we can make a better contribution to knowledge in usage-based linguistics if we spend more time with complex systems and less time trying to fit in with formal ideas incompatible with linguistics derived from language behavior.

Let us begin, briefly, with complex systems. In Melanie Mitchell's definition (2009: 13), a complex system is "a system in which large networks of components with no central control and simple rules of operation give rise to complex collective behavior, sophisticated information processing, and adaptation via learning or evolution."² Kretzschmar (2009) demonstrates how complex systems constitute speech in *The Linguistics of Speech*, focusing on four main points: "1) the continuum of linguistic behavior, 2) extensive (really massive) variation in all features at all times, 3) importance of regional/social proximity, and 4) differential frequency as a key factor in linguistic production both in regional/social groups and in collections of text corpora" (Kretzschmar 2009: 8). Complex systems are made up of massive numbers of components interacting with one another, and this "results in self-organization and emergent order" (2009: 253). For speech, the "components" of a complex system are all of the possible variant realizations of linguistic features as they are deployed by human agents, speakers and writers. The order that emerges in speech is simply the fact that our use of words and other linguistic features is significantly clustered in the local communities in which we actually communicate, or in social groups. We tend to talk like the people nearby, either physically near or socially near, or both. The same thing happens in the different texts we write and types of conversations we have; we tend to use the same linguistic tools that others do when we are writing or saying the same kind of thing.

The terminology of complex systems can sometimes be daunting, but we can observe two technical characteristics shown by complex systems that are easy to understand and use, non-linear distribution and scaling.³ When the variant types of any linguistic feature are graphed

according to their token frequency (Fig. 1), the chart exhibits a non-linear asymptotic hyperbolic curve (or A-curve), characterized by a small number of highly frequent responses and a much larger number of less-frequently-occurring responses (the long tail) among 109 variant terms for what people often call a *thunderstorm*.

Fig. 1 Atlas *Thunderstorm* data showing A-curve



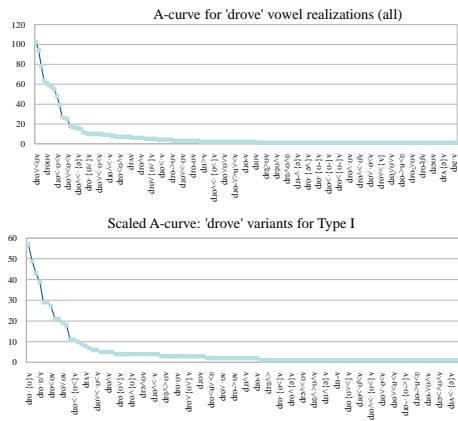
The concept of the A-curve will be familiar to those who know Zipf's Law, which states that the frequency of words in a text is inversely proportional to their rank.⁴ The A-curve is a perfect example of the non-linear distributional pattern characteristic of complex systems.⁵ Since the use of any variant by a speaker is influenced but not entirely determined by what the feature usually does, where it comes from, and what kind of people use it, group preferences among the variants are emergent and inherently unpredictable, but they always do emerge.

The property of scaling is another characteristic of complex systems.⁶ Scaling in speech takes the form of repeating non-linear distributions of variants for the data overall and for every subsample (Fig. 2). Here, in data recently prepared by my student Kristen Fredriksen, we see the frequencies of vowel realizations in *drove*: there are 162 different realizations of the

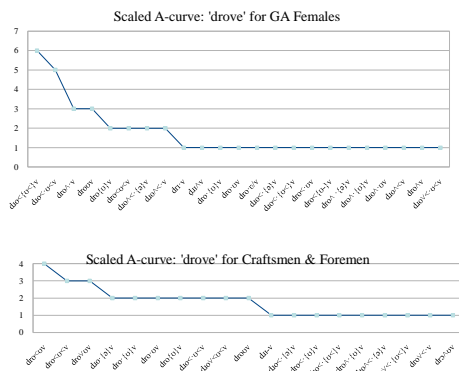
diphthong from all 1162 speakers in my survey, and 115 different realizations from the 582 folk speakers. As we go down in scale to fewer speakers, there are 23 different realizations from the 54 women interviewed in Georgia. There are only 5 different realizations from the 30 members of the occupational group "Craftsmen and Foremen," and yet we still see an indication of the non-linear pattern, particularly through the presence of the long-tail of occurrences that each happen only once or twice.⁷

Fig. 2 Vowel realizations in *drove* (Kristen Fredriksen)

Vowel realizations in *drove* (Kristen Fredriksen)

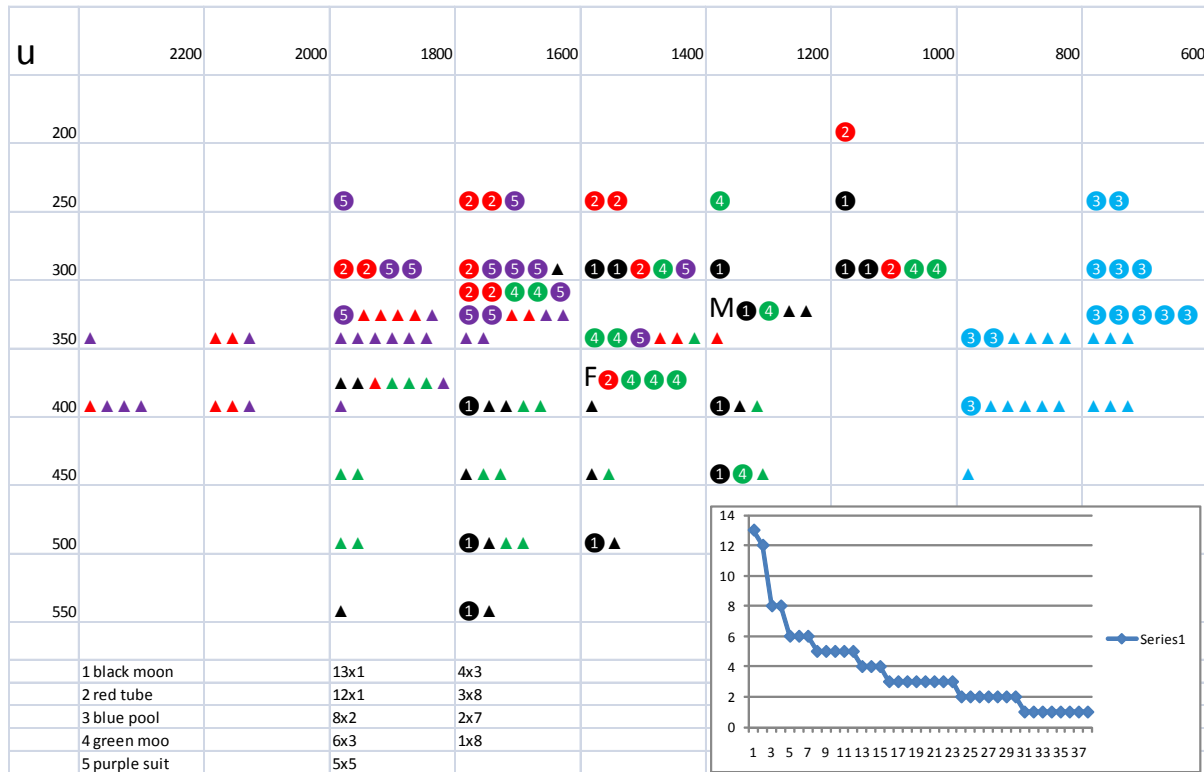


More vowel realizations in *drove* (Kristen Fredriksen)



We can also see this effect in acoustic phonetic data from a recent study of vowel realizations among 18 students at my university, in the number of tokens in each cell of a grid overlaid on acoustic F1/F2 space (Fig. 3; Kretzschmar, Dunn, and Kim 2011).

Fig. 3 Distribution of UGA /u/ realizations (148 tokens, 5 environments)



The frequency ranking overall is shown at bottom right, but clearly we could make separate and somewhat different plots for the men (the higher and backer circles) and the women (the lower and fronter triangles).⁸ Besides impressionistic and acoustic study of speech sounds, study of text corpora shows that the same kind of distributional patterns occur in different text types (Kretzschmar 2009: Ch. 5), and we will see a bit later that the same distributional pattern obtains for grammatical constructions.

We can see from these examples that these two properties of complex systems, non-linear distribution of frequencies of a large number of variants which we can chart at any level of scale, must interact in a complex, multidimensional way. Some subset of the same overall list of variants exists for every group of speakers we care to measure, and yet the frequency ranking of

those variants is likely to be different for each group we observe. One practical consequence of this complexity is that we will observe different results from the same data set at different levels of scale, and we can also expect to get different results from parallel samples taken from one group at different places or from different groups at one place. The top-ranked variant is much more common than the other variants for groups of any size, which means that we may well recognize the top-ranked feature as being particularly associated with its group of speakers. Thus the non-linear distributional pattern lets us tell groups of speakers apart. Although not demonstrated here, the same is true for different text types, including different kinds of conversations: we can expect to see different top-ranked variants for any linguistic feature in different text types, and different rank orderings, that allow us to tell the text types apart. This is how we know that some speaker or writer matches our expectations for who they are and what they are trying to say, or at least who and what they sound like on any given occasion. I would like to emphasize here that these two principles are not occasional--they apply to every linguistic feature, and so they are built into every aggregation of features, for whatever population of speakers, that we may ever wish to construct.

Now, let us go back to the early days for usage-based grammar and complex systems, to Paul Hopper and his influential 1987 paper, "Emergent Grammar." He did not derive his term "emergent" from complexity science, but his position accords exactly with complex systems:

This is, then, roughly the context in which the term Emergent Grammar is being proposed. The term 'emergent' itself I take from an essay by the cultural anthropologist James Clifford, but I have transferred it from its original context of 'culture' to that of 'grammar'. Clifford remarks that 'Culture is temporal, emergent, and disputed'.... I believe the same is true of grammar, which like speech itself must be viewed as a real-time, social phenomenon, and therefore is temporal; its structure is always deferred, always in a process but never arriving, and therefore emergent....

Hopper borrowed the term *emergence* from cultural anthropology, an example of how similar ideas can be "in the air" at a moment in time. The key point here is that Hopper maintains an idea of grammar as structure, but one that is never instantiated in the speech that generates it. If the structure of grammar is "always in a process but never arriving," a wonderful phrase, this

means that we can conceive of grammar as a rational object, but one that is always only indirectly related to speech. In another expressive piece of prose, Hopper explains further:

The notion of emergence is a pregnant one. It is not intended to be a standard sense of origins or genealogy, not a historical question of 'how' the grammar came to be the way it 'is', but instead it takes the adjective emergent seriously as a continual movement towards structure, a postponement or 'deferral' of structure, a view of structure as always provisional, always negotiable, and in fact as epiphenomenal, that is at least as much an effect as a cause.

Grammaticalization, as we now call this process, is not an explanation for current structure in language, but instead should be understood as "continual movement." Structure is "epiphenomenal" not because it is unimportant, but because it is always contingent and never directly observable. Finally, Hopper affirms that

Because grammar is always emergent but never present, it could be said that it never exists as such, but is always coming into being. There is, in other words, no 'grammar' but only 'grammaticization'- movements toward structure which are often characterizable in typical ways.

I would modify this striking statement only to clarify that, while grammar never exists as such in language in use, it can well exist as a description of regularities indirectly derived from speech performance by perceptual means. This, in fact, is just what all linguists do.

If we observe the distributional pattern of linguistic features at any moment in time, we can get more historical information than we might have thought. Paul Hopper used the example of the development of the English indefinite article *a/an*:

To take just these three functions of the predecessor of *a/an* in Old English, we find in modern English not a uniform, over-all weakening of the meaning, but rather a situation in which the weakened meanings and the older stronger meanings exist side-by side.

They introduced me to a young woman [whose name was Ethel]; My husband and I went to a showroom to pick up a new car we had ordered. After we had test-driven it, the salesman asked us ...

VS.

Birds of a feather flock together; They are all of a kind; A stitch in time saves nine; A penny saved is a penny earned; How much is that picture-frame? - A dollar; Linguists of a theory attend the same conferences; What was left of the woods after they built the parking-lot? - A tree.

In other words, English has preserved the historical functions of *a/an*, as well as developing new ones. As Hopper admitted, his retentions of archaisms are illustrated from proverbial language. Still, whether in proverbial language or in special domains like law or religion, we do still employ many such old constructions, which gives them a low frequency on the modern A-curve. The indefinite article is not an exception but a good example of a general capacity of speech to retain old feature variants at low frequencies. Stockwell and Minkova, for instance, reported in 1988 that it was possible to find vowels at every stage of the Great Vowel Shift in modern use somewhere in the Survey of English Dialects. Similarly, Allison Burkette (2001) has shown that in the modern A-curve of different words used for *bureau*, as constructed from our survey of American English, low frequency variants in the long tail are often names for that piece of furniture used in Colonial pattern books for furniture makers. In historical semantics, one of the points made by C. S. Lewis in his *Studies in Words* (1960) is that the meanings of key words in our culture have "ramifications." That is, words do retain their older meanings even while they gather new ones.⁹ Thus, for grammatical constructions, vowels, the lexicon, and semantics, all of the evidence suggests that historical forms tend to be retained as low-frequency variants in the tail of the non-linear distribution of contemporary usage. By the same token, the long tail is where we will find novel features as they enter the language. For Paul Hopper in 1987, then, grammaticalization is a process in which particular variants are not selected as being "grammatical," but instead just become more frequent at the expense of older variants, and then themselves become less frequent as newer variants arise and increase in use.

The next stop on our tour again features Paul Hopper, this time with Elizabeth Traugott in their 1993 book *Grammaticalization*. They define their key term as follows (xv):

We define grammaticalization as the process whereby lexical items and constructions come in certain linguistic contexts to serve grammatical functions, and, once grammaticalized, continue to develop new grammatical functions. It is the process whereby the properties that distinguish sentences from vocabulary come into being diachronically or are organized synchronically.

This statement cannot be more different from the 1987 article. Yes, grammaticalization is a process, but it is defined by how the properties of sentences "come into being" or have synchronic organization. Grammar has become an object. Further to this point they say (1):

Grammaticalization . . . refers to that part of the study of language that focuses on how grammatical forms and constructions arise, how they are used, and how they shape the language. The framework of grammaticalization is concerned with the question of whether boundaries between categories are discrete, and with the interdependence of structure and use, of the fixed and the less fixed in language. It therefore highlights the tension between relatively unconstrained lexical structure and more constrained syntactic, morphosyntactic, and morphological structure. It provides the conceptual context for a principled account of the relative indeterminacy in language and of the basic non-discreteness of categories.

So, grammaticalization is a process that exists with reference to a framework, a set of categories which are non-discrete. Items may be "more" or "less" grammatical, i.e. more or less a part of the framework. Grammar is a framework which is more or less contingent (because the categories are non-discrete, not naturally given as discrete), but which is nonetheless objectified, reified, and above all characterized by constraints of "syntactic, morphosyntactic, and morphological structure." This reification can also be observed in the use elsewhere by the Hopper and Traugott of nominalizations like "reanalysis," and related passive constructions like "have been generalized, or analogized," which leave out the agent for the action. The implication of such usage is that users of language are the agents, but really the agents who are doing the reanalyzing, generalizing, and analogizing are the linguists themselves, as they attempt to make useful idealizations from speech data.¹⁰ Grammatical categorizations have become the target of the argument, in that the process shows the non-discreteness of grammatical categories. What may have begun as a usage-based argument has now become highly formal, so that Hopper and Traugott treat the hypothesis of unidirectionality cross-linguistically and make it a "metacondition on how particular grammatical constructions will change" (95).¹¹ So, in *Grammaticalization* the book, what had started in "Emergent Grammar" the article has become radically different, radically more aligned with formal linguistics, and in so doing the most

brilliant insight of Hopper's article, the idea of grammar as continual movement, has been returned to a much more mainstream formal discussion in the book.

Another stop on the tour, Joan Bybee's 2001 *Phonology and Language Use*, cites Hopper's article but does not cite Hopper and Traugott. Nonetheless, she describes grammaticalization much as it is presented in the earlier book. The first and last lines tell the story. (2):

The present work proposes to demonstrate that the focus on structure needs to be supplemented with a perspective that includes more than just structure, a view that includes two other important aspects of the language phenomenon--the material content or substance of language, and language use. The SUBSTANCE of language refers to the two polar ends--phonetics and semantics--that language molds and structures, the two ends between which language forms the bridge. Language USE includes not just the processing of language, but all the social and interactional uses to which language is put. For present purposes, in the context of phonology, the frequency with which certain words, phrases, or patterns, are used will be shown to have an impact on phonological structure.

Bybee's focus remains on abstract structure, which she will "supplement" with information from phonetics, semantics, and language use. The latter, in the form of the relative frequency of particular words, phrases, or patterns, will merely have an "impact" on phonological structure. Thus phonological structure is a given, an assumption, that usage can only enhance or affect. Bybee (19) specifically invokes what she describes as "the goal of phonology as conceived by generative theory" (1) relations among similar (but different) sounds taken to be the same (allophones); 2) relations among variants of morphemes; 3) phonological units; and 4) language specific and universal properties of relations and units), and then claims that her

alternative model, which does not make the same assumptions as generative phonology, can accomplish these same goals as well as accommodate other facts about phonological structure....

Not only does Bybee take structure as a given, she aligns her discussion exactly with the goals of generative phonology. From such a beginning, it is difficult to imagine how much enhancement, or what sort of impact, language in use can actually have in a usage-based linguistics.

Bybee's University of New Mexico is not far from the Santa Fe Institute, and so it is fitting that she addresses complex systems:

This book is a linguist's book: it applies the established methods and data of linguists to the understanding of language as an emergent system resulting from the general cognitive capacities of humans interacting with language substance over many instances of language use" (18)

The key terms of complexity theory are here: emergence, speakers as agents involved in a great number of interactions. Bybee wants to see "segments as emergent units" (85). Her argument proceeds from the connectionist cognitive network model:

In network models, internal structure is emergent--it is based on the network of connections built up among stored units. The stored units are pronounceable linguistic forms--words or phrases stored as clusters of surface variants organized into clusters of related words. ... Units such as syllables and segments emerge from the inherent nature of the organization of gestures for articulation.

Lets unpack this statement. Bybee suggests that items from the stream of speech, pronounceable linguistic forms, are retained in memory, and that the storage of many items leads to "clusters" at a higher level of organization. Then further, Bybee claims that (86)

The view that phonological representations are self-organizing means that units of analysis, such as segments and syllables, are emergent units and are permitted to have gradient properties. This view does not insist upon one unit of uniform size for describing all speech, but rather proposes that the organization of linguistic material into units depends entirely upon the substantive properties of that material.

Self-organization, another key term from complex systems, thus leads in Bybee's view to the instantiation in the brain of linguistic structure, units like segments and syllables that linguists manipulate. The problem is that her a priori assumption of linguistic structure means that Bybee has prejudged the outcome of emergence. As we have seen, every individual is at the nexus of many groups of people, regional and social, and each one of those groups will have its own distributional pattern of many variants for any feature we name. If we revisit revisit Fig. 3 showing variant realizations of the /u/ vowel in a group of young Georgia speakers, we can see not only that the men (circles) are different from the women (triangles), but that the different environments for the sound (red and purple before stops, green and black before nasal or juncture, and blue before -l) each have their own distributions. Yes, we can create the overall distributional chart for acoustic realization for the phoneme /u/ in this group of speakers, as

shown at the lower right, but there is no reason to privilege the units that linguists normally talk about, like phonemes, which derive essentially from the simplifying generalization of a formal model; we could choose to emphasize the different environments instead. Moreover, when Bybee talks about frequency, she most often refers to Francis and Kučera's *Frequency Analysis of English Usage* (1982), derived from the Brown corpus of written material from the 1960s. Her leap to top-level analysis of English conceals the kind of variation on the ground that we have seen for the Georgia /u/ vowel. Her argument about the emergence of cognitive structure in individuals cannot fairly be related to the Brown Corpus, since no individual, not even Francis or Kučera, has ever been exposed to speech at the same frequencies of variants as occur in the corpus. Bybee's a priori assumption of formal structure leads her to take the complexity out of the complex system.

Next stop, Goldberg's 2006 *Constructions at Work*, which advances from her 1995 book, *Constructions: A Construction Grammar Approach to Argument Structure*. While it brings to bear another decade of work in the field, her 2006 book maintains the key point of her earlier work, that "all levels of grammatical analysis involve constructions: learned pairings of form with semantic or discourse function, including morphemes or words, idioms, partially lexically filled and fully general phrasal patterns" (5). Her other essential point is that "Constructions are combined freely to form actual expressions as long as they are not in conflict" (10). This of course differs sharply from formal models by rejecting linguistic nativism and biologically-coded universals, and also by favoring an additive, unitized grammar of constructions as opposed to a system of rules which derives surface structure from underlying levels in the rule system. Still, as we saw in Bybee's work, Goldberg does value continuity with formal linguistics even as she attacks it on many fronts, as shown here (4):

Constructionist approaches share certain foundational ideas with the mainstream "generative" approach that has held sway for the past several decades (Chomsky 1957, 1965, 1981). Both approaches agree that it is essential to consider language as a cognitive (mental) system; both approaches acknowledge that there must be a way to combine

structures to create novel utterances, and both approaches recognize that a non-trivial theory of language learning is needed.

Language is still a system instantiated in the mind. Goldberg's focus is still on structures. In saying this I in no way wish to minimize the departure of the constructionist model from the generative model, but as with Bybee, to make these assumptions prejudices the outcome of what usage may do. Goldberg's subtitle for *Constructions at Work* is "The Nature of Linguistic Generalizations." She has assumed that this nature will be structural. Thus, Goldberg can say that (45)

constructionist approaches are generally USAGE-BASED: facts about the actual use of linguistic expressions such as frequencies and individual patterns that are fully compositional are recorded alongside more traditional linguistic generalizations.

"Traditional linguistic generalizations" will thus be "supplemented," to use Bybee's term, by information derived from usage. Also like Bybee, Goldberg generally confines her analysis to the top-level, to "English" or to "Standard English," usually without comment. Thus, regarding the "took off screaming" construction, she can say that it "appears to be a serial verb construction of English, despite the fact that English does not allow serial verbs in general" (52), or that "the data reveals a family of closely-related constructions in English" (52), or that "while the ["take off screaming"] construction is part of Standard English, the ["go tell it on the mountain"] construction is restricted to informal speech" (53).¹² Goldberg values continuity with formal traditions in linguistics and accepts top-level, whole language analysis as the appropriate way to talk about usage. In so doing, Goldberg's assumptions and level of analysis thoroughly condition what findings she can achieve.

On the other hand, Goldberg's treatment of frequency effects does advance beyond Bybee's. In a section on "Skewed input" Goldberg reports on a corpus of transcribed mother-child conversations that (75-76)

In analyzing the mothers' speech we found a strong tendency for there to be one verb occurring with very high frequency in comparison to other verbs used in each of the constructions analyzed. . . . For example, *go* accounts for a full 39 per cent of the uses of the intransitive motion construction This high percentage is remarkable since this

construction is used with a total of thirty-nine different verbs in the mothers' speech in the corpus.

Besides *go* in the intransitive motion construction, Goldberg also cites *put* as occurring 38% of the time out of the forty-three verbs used in the transitive construction with motion, and cites *give* as occurring 20% of the time out of the thirteen verbs used in the ditransitive construction. This report of course matches what we should expect of non-linear distributions in a complex system. Goldberg specifically mentions Zipf's work on skewed distributions of words in texts, but does not make the connection, here or elsewhere in the book, to complex systems. Indeed, she invokes methods used elsewhere in linguistics to make a "fair" evaluation of linguistic structure by using statistical means to limit the effect of high-frequency variants (76-77). Still, the combination of her specific reference corpus with recognition of the skewed input for the children's language learning allows Goldberg to make a general statement that, for the most part, is not tied to traditional formalisms (92):

One particular verb accounts for the lion's share of tokens of each argument frame considered in an extensive corpus study The dominance of a single verb in the construction facilitates the association of the meaning of the verb in the construction with the construction itself, allowing learners to get a "fix" on the construction's meaning. . . . In this way, grammatical constructions may arise developmentally as generalizations over lexical items in particular patterns.

Only the last sentence invokes the assumption of traditional linguistic practice, by locating "grammatical constructions" as the outcome of making "generalizations over lexical items in particular patterns." Otherwise, the statement captures a genuine and important distributional effect that arises from a particular group of speakers, and connects it to individual cognitive development.

Michael Tomasello's 2003 *Constructing a Language* was published three years before Goldberg, but in many ways Tomasello advances further towards a more usage-oriented model. However, he begins not far from the previous books on our tour (5):

... in contrast to generative grammar and other formal approaches, in usage-based approaches the grammatical dimension of language is a product of a set of historical and ontogenetic processes referred to collectively as *grammaticalization*. When human

beings use symbols to communicate with one another, stringing them together into sequences, patterns of use emerge and become consolidated into grammatical constructions--for example, the English passive construction, noun phrase construction, or *-ed* past tense construction. As opposed to conceiving linguistic rules as algebraic procedures for combining words and morphemes that do not themselves contribute to meaning, this approach conceives linguistic constructions as themselves meaningful linguistic symbols--since they are nothing other than the patterns in which meaningful linguistic symbols are used in communication (for example, the passive construction is used to communicate about an entity to which something happens).

Again, grammaticalization is not here what it was for Hopper in 1987, a process that never creates a product, but instead a process that leads to "the grammatical dimension of language" which consists of patterns of use that "become consolidated into grammatical constructions" like the passive or noun phrases. Still, the idea of "constructions" is much more amenable to a usage-based approach than the "representational innateness" of Chomsky's biological approach, as Tomasello argues throughout the book.¹³ Instead, Tomasello recasts grammar as an inventory of constructions (5-6):

competence with a natural language consists of the mastery of all its items and structures, and these constitute a more complex and diverse set of linguistic representations than the "core grammar" of formal approaches. They include the highly canonical (core), the highly idiosyncratic (periphery), and many things in between.... A plausible way to think of mature linguistic competence, then, is as a structured inventory of constructions, some of which are similar to many others and so reside in a more core-like center, and others of which connect to very few other constructions (and in different ways) and so reside more toward the periphery.

Gone here is the assumption that "the properties that distinguish sentences from vocabulary" should constitute the basic model for grammar for Hopper and Traugott. Gone too is the abstract structure of relationships that Bybee accepted from the generative model. Goldberg's strictly synchronic orientation (she does not mention grammaticalization) has been replaced with a wider view in which the present inventory of an individual can be located with respect to "historical and ontogenetic processes." Tomasello's "structured inventory of constructions" explicitly allows for the common and uncommon variants that we know that we will find for every feature of language in any group of speakers we wish to study.¹⁴ This view of history accords well with what we have seen from Hopper and Traugott.

Whereas Goldberg addresses usage and frequency mostly at the top level of language in her discussion of generalization, Tomasello makes usage in particular learning environments into a very concrete factor for language acquisition.¹⁵ His skill of "intention reading" (sharing attention with others, following attention and gesturing of others, ability to learn the intentional actions of others) is an apt characterization of what children learn and what speakers do all the time: it describes the "joint attentional frame"¹⁶ that speakers must have in order for feedback to occur, which over time allows for the emergence of speech patterns both for cognitively for individuals and socially in groups of speakers. Although he does not cite J. R. Firth, who first emphasized the importance of recurrent situations of language in use,¹⁷ Tomasello gives recurrence a central role (88):

If a child were born into a world in which the same event never recurred, the same object never appeared twice, and the adult never used the same language in the same context, it is difficult to see how that child--no matter her cognitive capabilities--could acquire a natural language.

Recurrent situations of use--objects, events and context--are here explicitly joined to the "same language." The frequency of recurrence thus comes to the fore, but as for Goldberg, not complex systems, which are not mentioned in the book.¹⁸ Tomasello suggests that frequency is not a monolithic factor in acquisition (79):

... it is not the case that the more often a child hears a particular word, the quicker she learns it. As just two examples, the words *the* and *a* are among the most frequent words that young children hear, but they do not learn these words particularly early.

This observation crucially invokes the importance of the the process of language acquisition, not just frequency, in cognitive development.

It is well worth expanding on this point. Tomasello makes the general statement that "across many of the world's languages children initially seem to learn adult nouns more readily than adult verbs and other types of words" (45). In order to learn words at all, children must be able to recognize them in the stream of speech, and to associate them with meaning within their experience. The usual explanation for early learning of nouns and names is that these units are

the easiest to associate with concrete aspects of experience (46-50). As this Fig. 4 shows, the type frequency of nouns and social terms like names is almost 90% of a child's earliest vocabulary, with verbs and other words together near 10%.

Fig. 4 Type frequency, early vocabulary

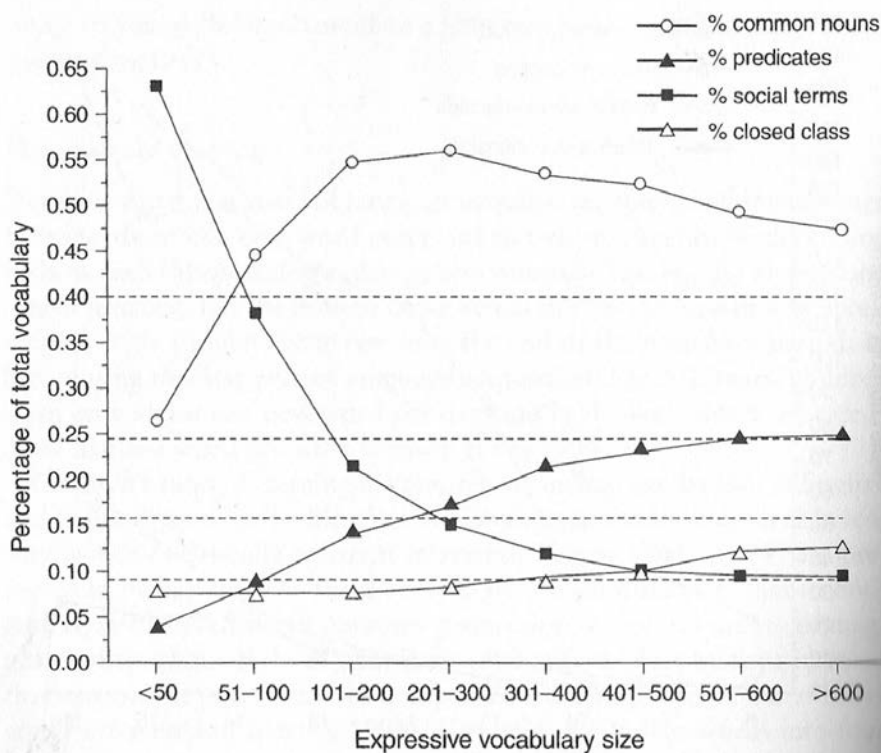


Figure 3.3. Vocabulary composition for Italian children from 18-30 months. From Caselli et al. (1995); reprinted with the permission of Elsevier Science.

As the vocabulary grows, social terms greatly decrease while nouns remain as a large percentage. However, by age 2 token frequency begins to assert itself (Fig. 5); in Tomasello's example list of the 25 most frequent words for one two-year-old, there are no nouns, and only one name-- "Mummy."

Fig. 5 Most frequent words for a two year old

Table 3.1 One child's 25 most frequently used words at age 2;0.

<i>Reference</i>	<i>Movement and location</i>
it	where's
one	here
that	there
the	on
a	down
this	go
it's	
that's	<i>Desire and prohibition</i>
my	want
	more
<i>People</i>	don't
I	
Mummy	<i>Conversational devices</i>
you	yeah
	no
	oh
	and

Tomasello explains that (81-82):

...these 25 words, which represent less than 2 percent of this child's vocabulary, account for 45 percent of all the words she says in terms of token frequency. What this list and these facts demonstrate is that as soon as just a small amount of grammar begins, young children's utterances are peppered with a relatively small number of high-frequency lexical items such as certain pronouns and function words with highly recurrent discourse functions--with the more well-known nouns and verbs, which are typically thought of as the prototypical items in young children's vocabularies, used relatively infrequently as their specific referents occur in the child's experience at only irregular intervals.

Thus learning begins with isolates (nouns, names), and then non-linear frequency takes over.

"Highly recurrent discourse functions" account for the distributional pattern according to

Tomasello. *A* and *the* are on the list now for the two-year-old because they recur with a great

many nouns, while the situation for the use of any particular noun will recur only seldom.¹⁹ To

see the non-linear pattern indicates not merely the fact that the child has "just a small amount of

grammar," but that the child has now passed the isolate stage and has achieved the threshold for

normal linguistic interaction.²⁰ These facts show that grammar is developmental, and of course

the child's command of her language will continue to develop. Indeed, as Tomasello later observes about the acquisition of nominal constructions (198):

Children learn and use nominals in their very earliest holophrastic language, but true skill with nominals comes very slowly because successfully identifying referents for different listeners in different discourse contexts requires sophisticated skills of perspective-taking (theory of mind).

A good example of this point is Hopper's high and low frequency uses of the determiners *a* and *the*, which would take learners some time to experience and acquire. Tomasello's phrase "true skill" is a good one, in that it does not mean that children acquire any fixed grammar of nominals, but instead implies that the developmental process of skill with nominals can continue for a lifetime.

The interaction of recurrent situations for language use, frequency effects, and children's developmental process of language acquisition are again illustrated nicely by Tomasello with charts of the relative frequency of construction types by mothers and by children (Fig. 6).

Fig. 6 Mothers' and children's construction frequencies by type

112 Constructing a Language

Table 4.1 Most general construction types mothers use in talking to their 2-year-old children.

	Current study		Wells (1983)	
<i>Fragments</i>		.20		.27
one word	.07		.08	
multi-word	.14		.19	
<i>Questions</i>		.32		.22
wh-	.16		.08	
yes/no	.15		.13	
<i>Imperatives</i>		.09		.14
<i>Copulas</i>		.15		.15
<i>Subject-Predicate</i>		.18		.18
transitives	.10		.09	
intransitives	.03		.02	
other	.05		.07	
<i>Complex</i>		.06		.05

Source: Cameron-Faulkner et al. (in press).

Table 5.1 Adult categorization of utterances 2-year-old children use in talking to their mothers (these are not child constructions).

<i>Fragments</i>		.69	N/NP (.39) = my finger
one word	.28		V/VP (.16) = nearly fell over
multi-word	.42		PP (.04) = on the blanket
<i>Questions</i>		.04	Wh = Where that go?
wh-	.04		Y/N = Are you writing?
yes/no	.003		
<i>Imperatives</i>		.03	Open it now.
<i>Copulas</i>		.07	That's Grandpa.
<i>Subject-Predicate</i>		.16	Tr = Mummy build a tower.
transitives	.09		Intr = I trip.
intransitives	.04		Oth = It make me sick.
other	.03		
<i>Complex</i>		.002	I want you sit there too.

Source: Compiled by Thea Cameron-Faulkner.

On the childrens' side, we see that the frequency data does not match what their mothers say. The children have many more fragments and fewer questions, predictably almost no complex constructions yet, though perhaps surprisingly about the same percentage of subject-predicate constructions as their mothers. The children are certainly in the process of developing their language, but another reason that they are not exactly imitating their mothers is because they are talking like children: they are interacting in recurrent situations for discourse that are asymmetrical with adults. On the mothers' side two experiments are represented, and the frequencies of constructions are not the same. The differences can be quite large, such as Wells group with a third more fragments, a third fewer questions, and half again as many imperatives. The finding is just what a complex systems view would predict, that different groups of speakers will have different distributional patterns for the variants of linguistic features we track (here, constructions). The consequence of the finding for children acquiring language is, as Tomasello points out (110): "children learn what they hear, and different children hear different things in different quantities." Tomasello uses this data to argue that "language acquisition is not just

triggered by the linguistic environment, as proposed by generative grammarians, but rather the linguistic environment provides the raw materials out of which young children construct their linguistic inventories" (110).²¹ Recurrence, frequency, and cognitive development may be hard to tease apart, but they all certainly appear in the data.

The last stop on our tour is Nick Ellis and Diane Larsen-Freeman's 2009 *Language as a Complex Adaptive System*. The book begins with a position paper by the members of the "Five Graces Group," which includes Joan Bybee, William Croft (another major figure in usage-based linguistics whose interests are not far from Goldberg's) and several others as well as the volume editors.²² With this roster of participants, complex systems has thus now become a focal mode for usage-based linguistics. They propose (2009: 5)

a usage-based theory of grammar in which the cognitive organization of language is based directly on experience with language. Rather than being an abstract set of rules or structures that are only indirectly related to experience with language, we see grammar as a network built up from the categorized instances of language use The basic units of grammar are constructions, which are direct form-meaning pairings that range from the very specific (words or idioms) to the more general (passive construction, ditransitive construction), and from very small units (words with affixes, *walked*) to clause-level or even discourse-level units.

The Five Graces continue:

Because grammar is based on usage, it contains many details of cooccurrence as well as a record of the probabilities of occurrence and co-occurrence. The evidence for the impact of usage on cognitive organization includes the fact that language users are aware of specific instances of constructions that are conventionalized and the multiple ways in which frequency of use has an impact on structure. (ibid)

The Five Graces clearly rely on terms and ideas that we have seen in usage-based linguistics.

Constructions are the basic unit, the components in the complex system. The Five Graces' assertion of a "network built up from categorized instances of language use" transforms

Tomasello's "inventory of constructions" with a term from complex systems, "network." They also apply grammaticalization in their position (6):

Historical changes in language point toward a model in which patterns of co-occurrence must be taken into account. In sum, "items that are used together fuse together" For example, the English contracted forms (*I'm, they'll*) originate from the fusion of co-occurring forms Auxiliaries become bound to their more frequent collocates, namely

the preceding pronoun, even though such developments run counter to a traditional, syntactic constituent analysis.

They continue to specify the goal of their position (7):

In the usage-based framework, we are interested in emergent generalizations across languages, specific patterns of use as contributors to change and as indicators of linguistic representations, and the cognitive underpinnings of language processing and change.

Grammaticalization, then, for the Five Graces is the universal process that describes the operation of the complex adaptive system of speech. Their general statement on complex adaptive systems should also be familiar from my earlier discussion of complex systems (2):

language as a [complex adaptive system] of dynamic usage and its experience involves the following key features: (a) The system consists of multiple agents (the speakers in the speech community) interacting with one another. (b) The system is adaptive; that is, speakers' behavior is based on their past interactions, and current and past interactions together feed forward into future behavior. (c) A speaker's behavior is the consequence of competing factors ranging from perceptual mechanics to social motivations. (d) The structures of language emerge from interrelated patterns of experience, social interaction, and cognitive processes.

Complex systems, then, appear to be a perfect match for usage-based linguistics.

Unfortunately, the final statement here, that the "structures of language" are the product of emergence from the operation of the complex system, carries forward the problems I have noted for the previous works, in that the Five Graces prejudge the outcome that may emerge from the complex system.²³ In focusing on grammar as a kind of object, a "network," they convert the process that Hopper emphasized in "Emergent Grammar" into a state.²⁴ For any single linguistic feature, the Graces refer to the process of "selection" as the emergence of a preference for one state of the feature, one variant, over other possible variants. They associate change with long-term alteration in social practices, "which in the extreme case leads to the fixation of [new forms] and extinction of [old forms]":

changes in lifestyles lead to the rise and fall of words and constructions associated with those lifestyles (e.g., the rise of *cell* [phone] and the fall of *harquebus*). In the latter case, the social identity and the social contexts of interaction lead to the rise and fall of linguistic forms that are associated with various social values by speakers.

The Five Graces' preference for, or "fixation" of, a grammatical state takes the dynamic movement of the complex system and freezes it, so that one variant of a feature becomes "grammatical" in the sense of having been selected. The Five Graces Group, in my view, has been too eager to identify grammar directly with one aspect of complexity science, states and state space. In so doing, again in my view, they lose Hopper's sense of continual movement, and also lose the benefit of understanding speech as a complex system.²⁵ And, as I have shown throughout our tour of usage-based linguistics, this problem arises because, as Walker Percy might have said, the authors presented here have stayed too close to the Scholastics and have not acted enough like Galileo in their observations.

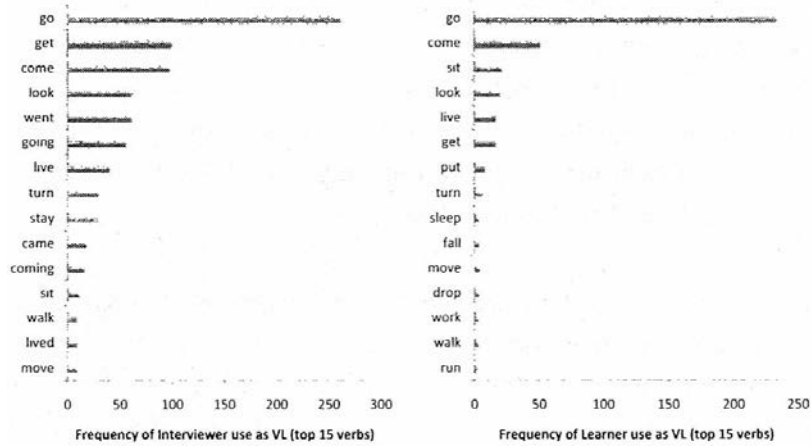
Still, there are many good things in the Ellis and Larsen-Freeman book. The position paper moves in the right direction, even if it carries some baggage. Complex systems is where usage-based linguistics ought to be going. We see in Fig. 7, for instance, in the Ellis and Larsen-Freeman article in the volume good examples of the non-linear distributional pattern in verb-argument constructions drawn from the speech of interviewers and learners in the European Science Foundation ESL corpus.

Fig. 7 Adapted from Ellis and Larsen-Freeman 2009: 101

VL constructions

Interviewer

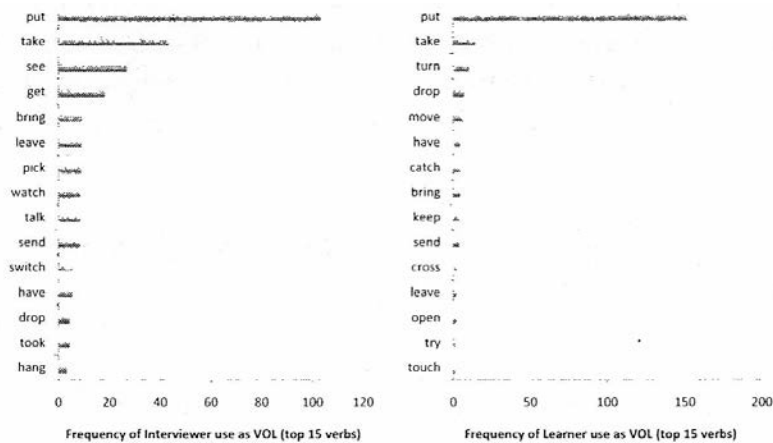
Learner



VOL constructions

Interviewer

Learner



Here, for the verb-locative construction (VL), the A-curve pattern of verb frequencies is clearly evident. *Go* is the most common verb by far in these constructions, with abruptly diminishing

frequencies for other verbs. Only the top 15 verbs are shown for both groups of speakers, so one can imagine a longer tail of uncommon verbs. Ellis and Larsen-Freeman invite us to "note the similar rankings of verbs across interviewers and learners" (101). Yes, *go* is at the top in both cases, and *come*, *look*, *get* are also common, but only 9 of the 15 verbs appear in both lists, and they are not in the same order. The same thing happens for the verb-object-locative constructions (VOL). *Put* is by far the most common verb for both groups, followed by *take*. However, this time only 7 verbs appear on both lists, and again the order is different after *put* and *take*. What is truly similar about these lists is the A-curve pattern, not the entire state of the network of nodes. The discovery of the non-linear pattern confirms Goldberg's earlier finding about the Zipfian nature of construction frequency rankings, now for ESL teachers and learners instead of mothers and children, an important extension of what we know.²⁶

So, at the end of the tour, what have we learned from usage-based linguistics and complex systems that can help us with analyses of aggregated data? First, all of the evidence indicates that we can expect non-linear distributions of variants for any linguistic feature we are interested to follow. This is true of speech sounds, whether measured impressionistically or acoustically, of words, and of constructions. As Hopper and Traugott and then Bybee have demonstrated, linguistic categories are not naturally given and discrete: over time expressions change in such a way that any boundaries for categories that we may wish to assert at one moment may well be breached a moment later. Still, we need to use categories so that we can count variant expressions and observe the non-linear distributions. The idea of "constructions" is perfect for this purpose, since the use of constructions does not entail a commitment to a fixed hierarchy of categories, and it may be applied to features as small as speech sounds and as large as discourse patterns. There is no reason that we cannot use traditional terms like "noun" or "verb" or "ditransitive" to name them, as long as we do not invoke an entire hierarchical system when we do so. The existence of non-linear distributions does not prove that any category we have selected is real or valid, however, because the scaling property of complex systems tells us

that we can expect to see a non-linear distribution anywhere we look, at any level, for any phenomenon, for any group of speakers. To say, as the Five Graces do, that "grammatical categories develop [through grammaticalization] in all languages" (8) puts the cart before the horse: frequency distributions occur for any constructions we decide to nominate, but we linguists are the ones who create the categories, not the operation of complex systems.

Grammaticalization really is just what Hopper said it was in 1987, a process of continual movement that is contingent on time, place, and circumstance and that does not allow grammar ever to be directly observable; it does not translate into categories for any state of a language. Neither does the non-linear distribution constitute evidence that there is a particular cause for the top-ranked variant to be where it is. The complex interaction of recurrence, frequency, and setting for language use rules out any simple cause for the state of a feature or language at any given time, except to say that the process of the complex system of speech always creates non-linear distributions.

Equally important, all of the evidence shows us that scaling is an unavoidable property of the complex system of speech. Tomasello's parallel data sets did not have the same frequency distributions, any more than my survey data had the same rankings at different levels of scale within the same data set. The verb-argument constructions described by Ellis and Larsen-Freeman are similar in pattern and top-ranked variants, but they are quite different in their ranking of verbs and inventory of higher-ranked verbs. The property of scaling has the unfortunate effect that we will find it difficult to generalize from one level of the language to any other. It is just not credible to jump back and forth between frequency evidence on a national level, like the Francis and Kučera rankings, and cognitive development in individuals. The operation of the complex system of speech in the cognitive development of children is essentially different in outcome from the operation of the complex system in a speech community at any level. As Tomasello has said, "children learn what they hear, and different children hear different things in different quantities" (110). We can never assume that others will share our

own experiences with constructions.²⁷ If we take care to match the assessments we make to the particular populations from which our data comes, we can make better generalizations, whether for a language as a whole, for national or regional varieties, for social groups, or for particular kinds of texts. The focus on the top-level of a language by early usage-based linguists, an overall grammar, is not a bad thing in itself but we cannot expect to connect work at that level with mothers and children or with individual cognitive development and expect our findings to match. Overall grammars are good for overall populations, but not for particular subgroups or individuals. The complex interaction of recurrence, frequency, and setting for language look different from every point of view at every scale of analysis.²⁸ And of course speakers, as agents using speech, are not dull bodies who have no choice in what features they want to use on any given occasion, including all the different kinds of conversations and text types that people enact, and so the complexity of speech is much, much greater than we have so far given it credit. The very idea that any single grammar could describe or generate such complexity is, frankly, beyond my imagination, because a complete single grammar for a language could never be motivated by detailed observation of speech production.²⁹ And of course the generative model and other formal models are not so motivated, and indeed were created to avoid the complications of language in use and to emphasize what might be the same about speech. Formal linguists have good reason to want to generalize away from observations of language production, and we can make good use of reasonably constructed abstract grammars.

We usage-based linguists also have good reason to want to observe language in use. When we are aware of the magnitude of the distributional problem, we certainly know that whatever experiment we conduct will never be enough to give us the kind of Big Picture answers that formal linguists have generated. Usage-based linguists should stop trying to align with formal linguists because formalists are trying to answer questions that essentially different from ours. Every experiment that we conduct, if it adequately describes the constructions it studies and the population of speakers who use them, makes another contribution to our knowledge of

the complex system of language in use. We may discover some crosslinguistic similarities, as usage-based linguists have already claimed to have done. Indeed, the logical conclusion we can draw from complex systems is that there is really only one language, the human language, and the phenomena that we have perceived as different languages are actually levels of scale within the overall complex system. To do a good job of language typology or cross-linguistic studies, we cannot ignore the complexity that we know to exist in aggregated data sets at any level of scale. The more evidence from well-controlled studies that we can build up, whether from small populations or from the sampling of large groups, the better picture we can develop of the way people use language at every scale of interaction. Nobody can do it all alone. Walker Percy actually needed more than just one Galileo to observe all the moons, planets, and stars in the linguistic cosmos. We should use his telescope together in order to build up a better understanding of language in use.

Abstract

In an influential 1987 paper, Paul Hopper described "emergent grammar" as "always in a process but never arriving." This wonderful phrase suggests that we can conceive of grammar as a rational object, a structure, but one which is never instantiated in the speech that generates it. Subsequently, Hopper became associated with "grammaticalization" (Hopper and Traugott 1993) and is cited regularly in new "usage-based" grammars such as construction grammar (e.g. Bybee 2001, Tomasello 2003, Goldberg 2006, Ellis and Larsen Freeman 2009). It turns out that complex systems, a growing movement in modern sciences such as ecology and genetics, is an excellent fit for "emergent grammar" as Hopper originally described it. Complex systems in language (Kretzschmar 2009) show non-linear frequency distributions of feature variants, possess scaling properties, and above all display emergent order. Unfortunately, usage-based grammars assume that the linguistic structure of a language should be identified with the order that emerges from complex systems, when in fact Hopper was right all along. This paper will consider the current usage-based literature, and show how one might use the non-linear distributions and the scaling property of complex systems to create grammars and typologies that more accurately reflect the emergence that comes from language in use.

References

- Burkette, Allison. 2001. The Story of Chester Drawers. *American Speech* 76: 139-157.
Bybee, Joan. 2001. *Phonology and Language Use*. Cambridge: Cambridge University Press.
Ellis, Nick, and Diane Larsen-Freeman. 2009. *Language as a Complex Adaptive System*. Oxford: Wiley-Blackwell.

- Ellis, Nick, and Diane Larsen-Freeman. 2009. Constructing a Second Language: Analyses and Computational Simulations of the Emergence of Linguistic Constructions from Usage. In Ellis and Larsen-Freeman 2009, 90-125.
- Firth, J. R. 1935. The Technique of Semantics. *Transactions of the Philological Society*, 36-72. Reprinted in Firth 1957: 7-33.
- Firth, J. R. 1957. *Papers in Linguistics 1934-1957*. London: Oxford University Press.
- Francis, W. Nelson and Henry Kučera. 1982. *Frequency Analysis of English Usage*. Boston: Houghton Mifflin.
- Goldberg, Adele. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Goldberg, Adele. 2006. *Constructions at Work*. Oxford: Oxford University Press.
- Goldberg, Adele, D. Casenhiser, and N. Sethuraman. 2004. Learning Argument Structure Generalizations. *Cognitive Linguistics* 14: 289-316.
- Guastello, S., and L. Liebovitch, Introduction to Nonlinear Dynamics and Complexity, in Guastello et al., eds., *Chaos and Complexity in Psychology* (Cambridge: CUP, 2009), 1-40.
- Hopper, Paul. 1987. Emergent Grammar. *Berkeley Linguistics Society* 13: 139-157. (viewed at <http://home.eserver.org/hopper/emergence.html>)
- Hopper, Paul, and Elizabeth Traugott. 1993. *Grammaticalization*. Cambridge: Cambridge University Press.
- Kretzschmar, William A., Jr. 2006. Review of Geoffrey Sampson, *Empirical Linguistics*. *Journal of English Linguistics* 34: 161-165.
- Kretzschmar, William A., Jr. 2009. *The Linguistics of Speech*. Cambridge: Cambridge University Press.
- Kretzschmar, William A., Jr., Josh Dunn, and MiRan Kim. 2011. Implicational Scaling in Southern Speech Features. Paper presented at ADS/LSA, Pittsburgh.
- Kretzschmar, William A., Jr., and Samantha Knoll. 2009. How Speech Communities Differ. Paper presented at SHEL, Banff.
- Kretzschmar, William A., Jr., and Edgar Schneider. 1996. *Introduction to Quantitative Analysis of Linguistic Survey Data: Atlas by the Numbers*. Thousand Oaks, CA: Sage Publications.
- Kretzschmar, William A., Jr., and Susan Tamasi. 2003. Distributional Foundations for a Theory of Language Change. *World Englishes* 22: 377-401.
- Langacker, Ronald. 1988. A Usage Based Model. In Brygida Rudzka-Ostyn, ed., *Topics in Cognitive Linguistics* (Amsterdam: John Benjamins), 127-161.
- Larsen-Freeman, Diane, and Lynne Cameron. 2007. *Complex Systems and Applied Linguistics*. Oxford: Oxford University Press.
- Lewis, C. S. 1990. *Studies in Words*. 2nd ed. Cambridge: Cambridge University Press.
- Lindblom, Bjorn, Peter MacNeilage, and Michael Studdert-Kennedy. 1984. Self-organizing Processes and the Explanation of Phonological Universals. In B. Butterworth, B. Comrie, and O. Dahl, eds., *Explanations for Language Universals* (New York: Mouton), 181-203.
- Mitchell, Melanie. 2009. *Complexity: A Guided Tour*. Oxford: Oxford University Press.
- Percy, Walker. 1977. "Questions They Never Asked Me," *Esquire* 88 (December): 170, 172, 184, 188, 190, 193-194.
- Stockwell, Robert, and Donka Minkova. 1988. The English Vowel Shift: Problems of Coherence and Explanation. In D. Kastovsky and G. Bauer, eds., *Luick Revisited: Papers Read at the Luick Symposium at Schloss Lichtenstein* (Tübingen: Gunter Narr), 355-394.
- Stubbs, Michael. 1996. *Text and Corpus Analysis*. Oxford: Blackwell.
- Tomasello, Michael. 2003. *Constructing a Language*. Cambridge: Harvard University Press.

¹ The reference to Galileo is not uncommon, in fact used by linguists on both sides of usage-based study (see Kretzschmar 2006). I am grateful to my student Calen Verbist for making me aware of Walker Percy's ideas about linguistics. I am also grateful to members of the Workshop on Cross-linguistic and language-internal variation in text and speech, Freiburg, 2011, for many useful comments. The works on usage-based linguistics treated in this essay come from the American tradition in this area; for the development of the "language in use" model in Britain, see Kretzschmar 2009: Chs. 1, 5.

² Mitchell (2009: 95) later admits that "there is not yet a single science of complexity but rather several different sciences of complexity with different notions of what complexity means." She offers no fewer than eight approaches to the definition (96-111), one of which, fractal dimensionality, is the one chiefly applied in Kretzschmar 2009 and also in this essay.

³ In other fields description of self-organization uses data collected in time series, and applies complex mathematical operations to generate "attractors" (see, e.g., Guastello and Liebovitch 2009). We have never used time series data for language variation studies, because unlike the binary "on/off" data from many sciences, our data is "rich," containing lots of information per data point, and so it is expensive to collect and analyze; it is impractical to go back and repeat observations frequently enough for time series.

⁴ The word "law" is actually a miscategorization, though; the distributions in my survey data are always non-linear but variable in proportions. Thus we can refer to the "80/20 Rule" as a practical rule of thumb, in which about 20% of the types account for 80% of the tokens, and 80% of the types in the long tail account for only 20% of the tokens.

⁵ The best explanation for why the A-curve distribution occurs in speech is that "feedback" between speakers, the positive or negative impressions that speakers get from massive numbers of linguistic interactions within their regional and social settings of use. Feedback results in the maintenance within the group of a large number of variants for any feature, and a strong preference for just one or a few of the variants.

⁶ Scaling in complex systems can also be called (confusingly) a "scale-free network," which refers to the components in a complex system and the fact that they do not relate to each other at just one scale.

⁷ The number of tokens and number of possible realizations for a feature in an experiment affect the shape of the curve. I will present an analysis of this problem at the 2011 Methods in dialectology meeting.

⁸ The different cues also show tendencies, *moon/moo* vs. *tube/suit* as well as *pool* (see below). Linguistic environment contributes to the variability in tokens of /u/, but environment alone does not explain the variation. The non-linear curve of frequency ranking could be seen clearly, given enough tokens and an appropriate choice of cell size, for each environment separately, just as we can observe it for the environments together.

⁹ This fact may be obscured for incautious readers by what Lewis called the "dangerous sense," the meaning that is so frequent in modern usage that we automatically think of it. Lewis could have had no idea of A-curves, but his dangerous sense is of course the top-ranked meaning on the modern non-linear curve of meanings for a word.

¹⁰ 17: "Do we need in our analyses to continue to "stop the film" and fix the grammar of a language as we investigate its structure, or do we need to view "grammar" as a provisional way-station in our search for the more general characteristics of language as a process of organizing cognitive and communicative content?"

This is the statement in Hopper and Traugott closest to Hopper 1987, but still "way-station" is a location at a moment, still an object.

¹¹ One of the key claims of grammaticalization is the unidirectional development of more-fully-realized forms into more compressed or generalized and more grammatical forms. Hopper and Traugott's opening example is what happened to *Let us*, in which an original second-person imperative construction with the verb *let* is first contracted, then the meaning of *let* is generalized, and then the contracted *s* loses its relation to *us* and just becomes the last sound in the new *lets*, a discourse particle, which in turn may be even further reduced by loss of its *t* sound. Hopper and Traugott routinely transfer examples like these from their item-based shapes, as here the sequence of representations that show what happens to *let us*, to clines consisting of grammatical categories, as here word > affix > phoneme. Indeed, the clines presented later in the book are not always accompanied by examples of sequences of fully realized representations, so that the grammatical categories have to shift for themselves (Ch. 5).

(let) us > *(let)'s* > *(let)s* = word > affix > phoneme
lexical item > morphology
lexical item in specific context > syntax > morphology
major category (> adjective/adverb) > minor category
relational noun > secondary adposition > primary adposition > agglutinative case affix > fusional case affix
full verb > (vector verb >) auxiliary > clitic > affix
discourse > syntax > morphology > morphophonemics > zero

¹² In another aspect of the same problem, Goldberg routinely employs the formal linguistic notational scheme of using starred and questioned forms, as here (56-57):

- (34) Pat helped her grandmother walk up the stairs.
*Pat aided her grandmother walk up the stairs.
(35) ??Pat helped her grandmother in walking up the stairs.
Pat aided her grandmother in walking up the stairs.

To me, all four of these expressions are acceptable. One could spend time recapitulating the usual arguments about the validity of grammaticality judgments, but the point here, again, is just that Goldberg is willing to make them in the way that linguists traditionally have done so.

¹³ Tomasello summarizes the differences in this way (2003:284-290):

No linguistic items or structures that are universal
No poverty of stimulus
Linking does not work
Parameters do not help
Continuity assumption is demonstrably false
Performance factors and maturation "are simply unprincipled fudge factors"
Lexical learning makes theory indistinguishable from other learning theories
Evidence for biological bases does not demonstrate representational innateness

¹⁴ However, it does not help to allow that constructions may be more or less grammatical, among the core of common constructions or on a periphery of uncommon ones, as Tomasello suggested. Whether variants happen to be common or not, speakers in the group under analysis are still using them to some degree, and the long history of uncommon forms that Hopper and others have demonstrated shows that even the uncommon features are grammatical, not just speech errors to be excluded from analysis. Grammaticality, as derived from formal linguistics, is a binary property like statistical significance. Probability, as in the non-linear distributions from complex systems, is continuous. It is a good thing to recognize that constructions may be more or less common, but that assigns them to a continuous domain, and takes them out of the binary domain of grammar.

¹⁵ One of Tomasello's main contributions specifies the terms of interaction for speakers. His primary interest is language acquisition, and so he frames his discussion in terms of mothers and children, but the

principles he suggests correspond to the interactional setting for all speakers. Children need to learn what Tomasello calls "intention reading," which consists of sharing attention with others, following attention and gesturing of others, and the ability to learn the intentional actions of others. These skills are related to Grice's pragmatics (12, 23, 28). Children also need to exercise the cognitive skill of "pattern finding," i.e. the ability to form perceptual and conceptual categories of "similar" objects and events, the ability to form sensory-motor schemas, the ability to perform statistically based distributional analyses on various kinds of perceptual and behavioral sequences, and the ability to create analogies. These skills are necessary for children to acquire the appropriate use of any and all linguistic symbols, including complex linguistic expressions and constructions--and for speakers to continue to use the language they have learned. Rather than associate cognitive development with any assumption of abstract linguistic structure as Bybee and Goldberg did, Tomasello's "pattern finding" merely specifies the tools to use and not the goal to attain.

¹⁶ Tomasello (24-25) provides an excellent example of an American buying a Hungarian train ticket without any shared language. This shows the power of the "joint attentional frame" and "referential event" (giving money, etc) in allowing a successful transaction.

¹⁷ Firth claimed that language is used "in a relatively determined context [where] you are not free just to say what you please" (Firth 1935: 66, 70-71; cited in Stubbs 1996: 41).

¹⁸ While frequency is important for Tomasello, he does not yet have a framework in which to use it (15):

Frequency plays a large role in this process [grammaticalization and syntacticization] as well, as only relatively frequently used expressions will become highly predictable--which accounts for the well-known principle that the more frequently a word is used in a language the shorter it tends to be (Zipf's Law). Frequency is also crucial because, as is well known, constructions that occur frequently are often irregular.

Tomasello chooses to cite "Zipf's Law" not for the main idea of a relationship between word rank and word frequency in texts, but for his subsidiary comment that frequent words tend to be short. He also cites the idea that high frequency protects irregular features, a point also emphasized by Bybee, but one that leads away from the regular non-linear distribution of variants for all linguistic features. This is not surprising, since the first thorough expansion of Zipf's Law as a consistent property of individual linguistic features outside of words in texts was Kretzschmar and Tamasi 2003; Goldberg's treatment of the distributional pattern in constructions stems from Goldberg, Casenhiser, and Sethuraman 2004.

¹⁹ His term "peppered" is perhaps unfortunate, if colorful, because it makes the non-linear distribution out to be something random or exotic, when in fact we know from Zipf and complex systems that it is the normal state for word frequency.

²⁰ Tomasello (93) cites Dabrowska as calling syntactic constructions simply "big words." In highly inflected or agglutinative languages, in fact they are.

²¹ It is more difficult to agree with his statement that "The fact that most adults end up with fairly similar (though not identical) linguistic inventories does not negate the obvious fact that early in development children can only learn what they are exposed to" (110), because in fact adult linguistic inventories differ much more from place to place and from group to group than we give them credit. Moreover, every adult is at the nexus of many groups of speakers, which makes the cognitive status of any speaker into a multidimensional problem and renders individual speakers inherently unpredictable. The idea that adults end up in the pretty much the same place is not a demonstrated fact, but instead just another manifestation of a traditional linguistic assumption, that adults end up with competence at the overall level of a language.

²² The book is a collection of papers originally published as a special issue of the journal *Language Learning*. Bybee, Croft, and Adele Goldberg are among the authors of the articles.

²³ Elsewhere, when they write about computational modeling of language, they say that "Clearly, speakers need to be equipped with a prescription for producing utterances that may vary between speakers (a grammar)" (13).

²⁴ In the language of complex systems, a network is a set of nodes, the elements in the complex system, without regard to the condition of each node. A "state" is the condition of all of the elements in the system at one moment in time, or of a single element. For a single traffic light (to borrow an example from Larsen-Freeman and Cameron 2007), there are three possible conditions, red, yellow, or green, that the light could have at any moment in time. In the language of complex systems, the "state space," the set of possible states in the system, of the single traffic light consists of those three states. If we had two traffic lights, the state space would have 9 states (3^2 : RR, RY, RG, YY, YR, YG, GG, GR, RY). Linguistic features are not so simple as a traffic light by many orders of magnitude,

²⁵ The five Graces do acknowledge that, "because this categorization is ongoing during language use, even adult grammars are not fixed and static but have the potential to change as experience changes" (7). Even this statement suggests that adult grammars can achieve a different state (because a grammar must be a cognitive state), not that they are characterized by Hopper's "continual movement."

²⁶ Ellis and Larsen-Freeman go on to argue that the non-linear distribution influences the ESL learning process by making the top-ranked variant "prototypical" for the construction. Such an argument follows from traditional linguistic practice and prior practice in usage-based linguistics by selecting the single top-ranked variant as more grammatical than the others. The interpretation may thus be less satisfactory than the examples, but again, this article and the others in the collection are moving in the right direction.

²⁷ We cannot assume that speech is like uniform paint on a wall, where we can tell the color of the entire wall by looking at any small part of it. We need to use valid randomized statistical sampling when we measure the frequencies of constructions in different, particular populations. We can take samples for very small populations, as sociolinguists now do for communities of practice: so, we can sample from just a small section of a wall, and tell what color the section is, say the part of the wall that differs in color from being behind a picture for many years. Alternatively, we can sample from an entire wall and tell what color the wall is, say the wall in a room with an accent color. And again, we can sample from the walls of an entire room and tell what color the room is, whether it is dark or light overall. The judgment we make about the room as a whole will not tell us about the accent wall or the lighter patch where the picture used to hang, and vice versa, the small patch or the single wall will not tell us about the room as a whole. In some cases they might—there might not have been a picture above the patch we sample, and all the walls may be the same color. But in the complex system of speech not only can we not rely on such uniformity, we know just from our own experience, no sampling required, that English and other languages are continuously variable around us.

²⁸ Finally, let us consider how good counting might be carried out for aggregated analyses. We cannot consider just the presence or absence of constructions; we need to consider also the pattern of variability in their frequency. If we have, say, 1000 constructions to track (about the number of cues in my survey data, but of course far fewer constructions than we could recognize and aggregate for any natural language), we should theoretically make 1000 charts of the distribution pattern of variants for the overall population, and then make equivalent charts for every subgroup we want to track in order to observe the differences in top-ranked variants and in the order of variants for each case. So, in practical terms, in Kretzschmar and Schneider 1996, for example, we tracked eight regional and social variables whose gradations created 52 different categories for people (14 occupations, 18 regional areas, 7 levels of education, and so on). That makes 52,000 distributions to consider even before we might want to try any multivariate analysis. There are 381,024 possible pairwise combinations of our cell variables, and so 381 million distributional patterns to consider for our set of 1000 features, even at the first multivariate step. Now we know why such a system is said to be complex. If we take the analysis down to the level of individual speakers, and so need to compare their individual distributional patterns to groups and to the

overall population, the complexity would be about three orders of magnitude higher for my survey of over 1000 people, thus nearly a trillion patterns to compare.

²⁹ Kretzschmar and Tamasi 2003 argues that the same is true for Labov's idea of coexistent systems, parallel grammars associated with social classes. If we do try to construct overall grammars from usage data they will necessarily be highly selective, and they will share the problem of trying to create national-level representative corpora--it is not possible to achieve adequate sampling of individuals and text types for the task.