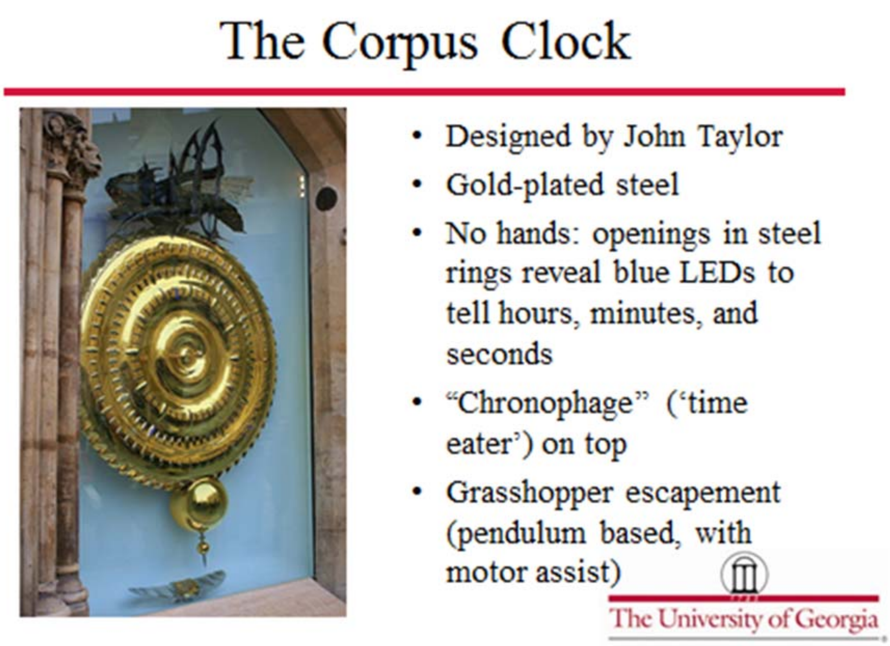
Complex Systems and the History of the English Language

Bill Kretzschmar, University of Georgia

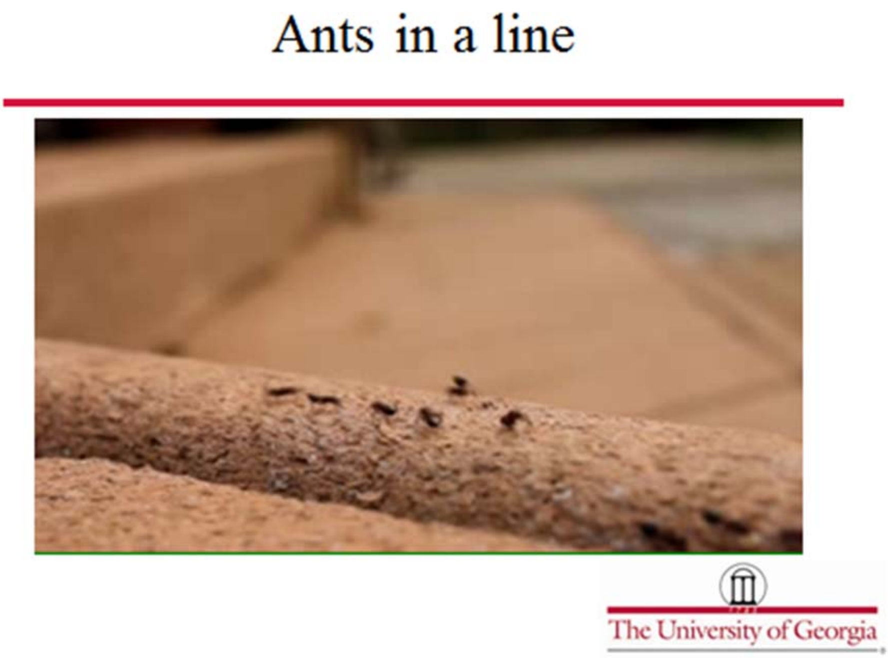
When I last visited Cambridge University, I found this in the Kings Parade.



Cambridge is one of those old places that seem never to change, and yet here was the new Corpus Clock right in the middle of things, on an outside corner of the staid Corpus Christi College (founded in 1352) and right across from the sixteenth-century Kings College Chapel, famous for its Christmas Eve choral broadcasts. Not long after my previous visit to the city in 2008, Stephen Hawking, who wrote *A Brief History of Time* (1988), had unveiled the clock and thus added a bit to time's narrative. The rather disturbing "chronophage," or 'time eater,' atop the clock eats the hours and the seconds as they pass. Its insect-like form was perhaps suggested by the name for the workings of the clock, a "grasshopper escapement" originally invented in the 18th century and here enhanced by an electric motor that winds a spring to keep the clock in motion. The other disconcerting thing about the clock, after its striking appearance and its equally amazing appearance at all in old Cambridge, is that it is only accurate every five minutes: time slows down for a while, even stops, and then speeds to catch up. Of course this trick is on purpose, something that, like the chronophage, is meant to unsettle us from our comfortable assumptions about time and its regular passage. Thus Stephen Hawking was the right person to unveil it, owing to his work in contemporary physics that challenges our assumptions about time and the universe around us, and thus too my thought to show it to you today, as a fitting emblem of the new ideas about our field of historical linguistics that I want to bring to you today.

The new science of complex systems (Mitchell 2009, Kretzschmar 2009), I would argue, is something that historical linguists not only *can* use but *should* use in order to improve the relationship between the speech we observe from historical settings and the generalizations we make from it. After a brief introduction to complex systems, I would like to take up the topic of language evolution from the point of view of the Corpus Clock, that the evolutionary process can speed up and slow down in time, rather than proceed in regular steps. I would then like to suggest some practical applications of complex systems to historical linguistics, things that we can do now to improve how we think about our work. My examples will come from the language whose history I know best, English, but at the same time it should be clear that the principles I will describe are not limited to English and can be applied to your own linguistic interest.

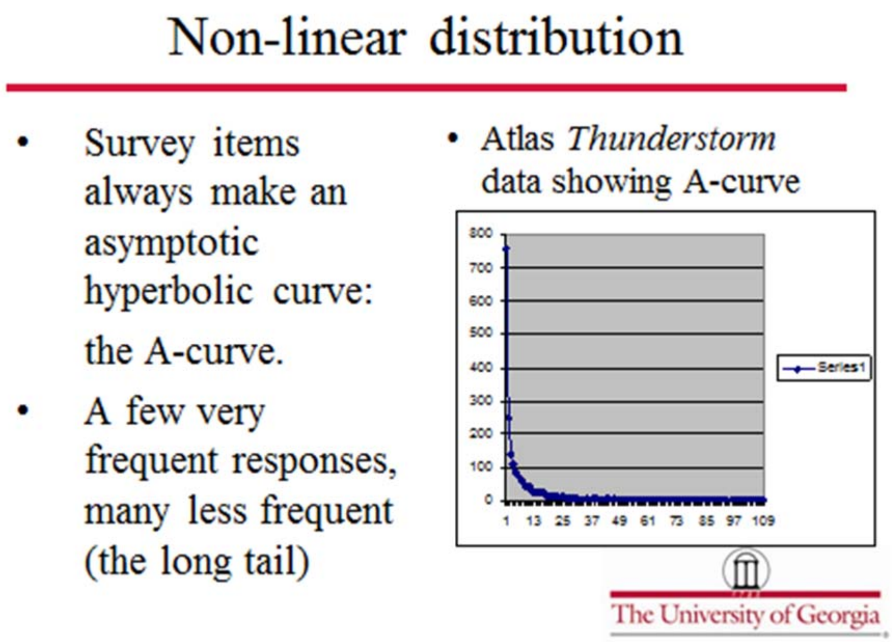
Complex systems, as described in physics, evolutionary biology, and many other sciences, are made up of massive numbers of components interacting with one another, and this results in self-organization and emergent order. Let's begin with a non-linguistic example and consider this line of ants (the ant section is derived from Mitchell 2009).



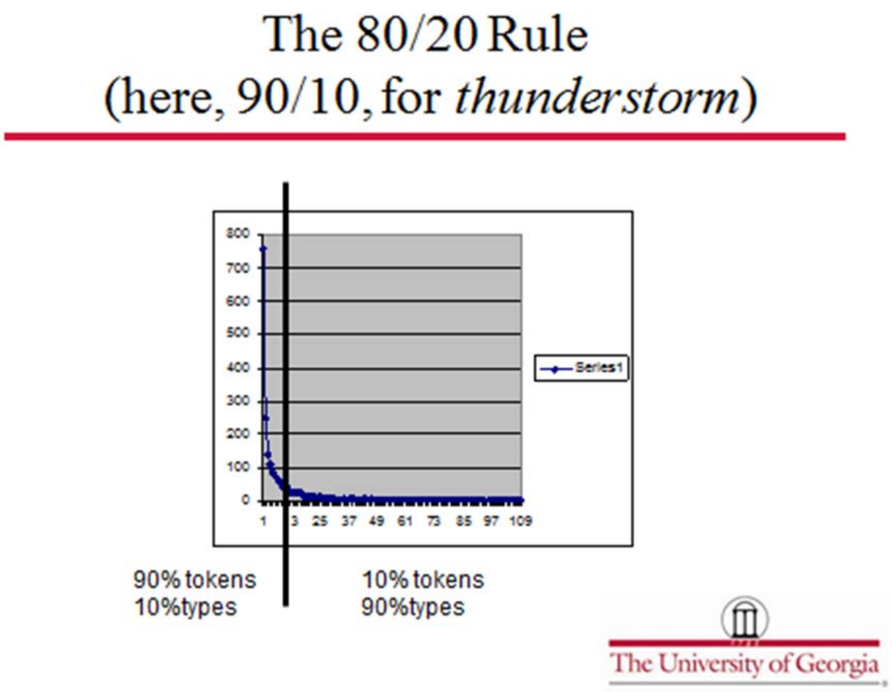
Ants are not smart. They can only do a few things, like exploit food sources, build nests, and defend themselves against intrusions, but no commissar of ants tells them to do any one of them. Instead, ants just happen to be doing one of these tasks at any given time. In searching for food, for example, ants wander around randomly; if they find some food, they leave chemical traces along their path to bring it back to the colony. Other ants can follow the traces, and leave more traces on their way back, till the path becomes a line like this one with lots of ants on it to exploit the food resource. However, not all of the ants in the nest follow the path; some keep foraging and some stay home on nest and defense duty, again not because they are told to do so by the queen or some ant general. When ants come boiling out of a nest when it is disturbed, they are not reacting to a chemical trace but to touching the antennae of other ants, as the slide sometimes shows, and then changing their behavior from food or nest duty to defense. However, not all of the ants leave to gather food or rush to defense given the stimulus to do so. Some stay on nest duty during a provocation, and some ants look randomly for food when a great many have joined the line to a known food source. What the ant does at any given time is influenced, but not determined, by what happens near it; when it detects chemical traces or touches antennae, an ant becomes more likely to enact one of the three behaviors. Individual ants are subject to feedback from other ants; in a primitive way, they exchange information. What looks to us like highly organized behavior is not controlled by any leader, and it is not absolutely determined by instinct or particular stimuli, but instead patterns of activity emerge from the instinctual behaviors of ants as they are conditioned by circumstances. These patterns of activity that emerge from the complex system make the whole more than just the sum of a few instinctive behaviors.

We can sum up the process of all complex systems in just a few principles: 1) random interaction of large numbers of components, 2) continuing activity in the system, 3) exchange of information with feedback, 4) reinforcement of behaviors, 5) emergence of stable patterns without central control. Complex systems like this were originally described in the physical and biological sciences, and the definite procedures of information exchange have been explored in computer science (e.g. Holland 1998). Stephen Hawking includes complex systems in his latest popular book about physics, *The Grand Design* (2010); and Stephen Jay Gould (2003) and many others have pursued complex systems in evolutionary biology and genetics. Complex systems also occur everywhere in speech, as I have described from first principles in *The Linguistics of Speech* (2009). For speech, the randomly interacting "components" are all of the different variant realizations of linguistic features as they are deployed by human agents, speakers. These "features" might be different pronunciations, or different words for the same thing, or different ways of saying or writing the same thing, really any aspect of speech that is recognizable for itself and therefore countable. The activity in the system consists of all of our conversation and writing. The exchange of information is not the same as sharing the meaningful content of what we say and write (which is exchange in a different sense), but instead our implicit comparison of the use of different components by different speakers and writers, as they use them in different kinds of conversations and writings. Feedback from exchange of information causes reinforcement, in that speakers and writers are more likely to employ particular components in future occurrences of particular circumstances for conversations and writing. Human agents, unlike ants, can think about and choose how to deploy linguistic variants, but that does not change the basic operation of feedback and reinforcement; we all make choices inside of the complex system of speech, in relation to current circumstances. The order that emerges in speech is simply the configuration of components, whether particular words, pronunciations, or constructions, that comes to occur in the local communities, regional and social, and in the occasions for speech and writing, text types and registers, in which we actually communicate. The process at work in complex systems just explains better what we already knew: we tend to talk like the people nearby, either physically near or socially near, or both, and we tend to use the same linguistic tools that others do when we are writing or saying the same kind of thing.

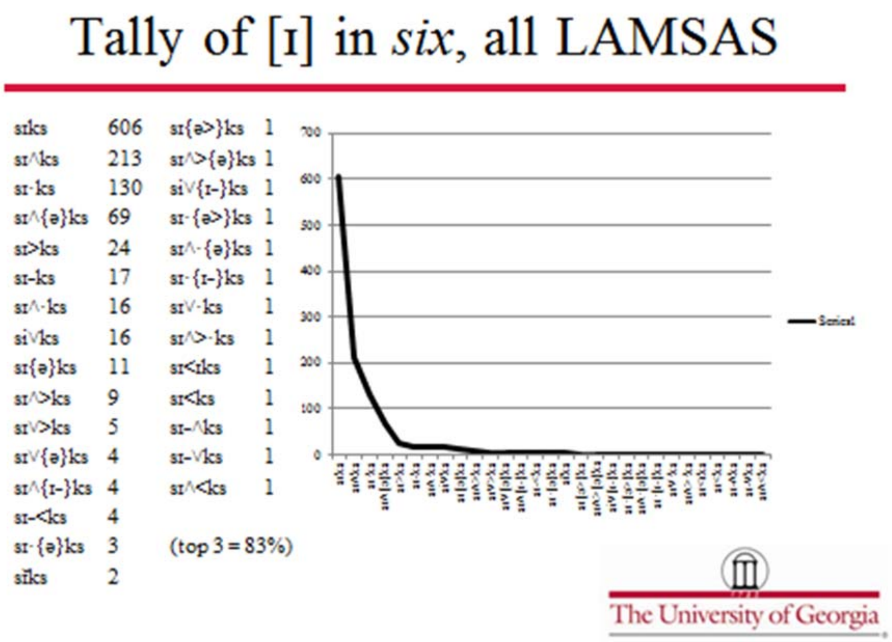
It is easy to see the patterns emerging from ant behavior, but it is much harder to watch emergence in speech over long stretches of time. Fortunately, the complex system of speech exhibits two technical characteristics that we can look for as signs that the system has operated: nonlinear distribution and scaling.



When the variant types of any linguistic feature are graphed according to their token frequency, the chart exhibits a nonlinear asymptotic hyperbolic curve (henceforth, A-curve), characterized by a small number of highly frequent responses and a much larger number of less-frequently-occurring responses (the long tail). The curve is the result of reinforcement from feedback in the continuing exchange of component variants for, here, what 1162 Americans in my LAMSAS survey of speech in the Eastern states call a thunderstorm. The concept of the A-curve will be familiar to those who know Zipf’s Law (1949), which states that the frequency of words in a text is inversely proportional to their rank. Jean Séguy noted the same nonlinear distribution of linguistic distance as a function of geography (1971), called Séguy's Law. Until now these "laws" have been viewed as curiosities by most linguists, but in complex systems the nonlinear distributional pattern occurs literally everywhere, for every feature in every survey I know of, in whatever language is being studied (English, German, French, Polish, Thai, and Japanese, from those my own students have studied so far).



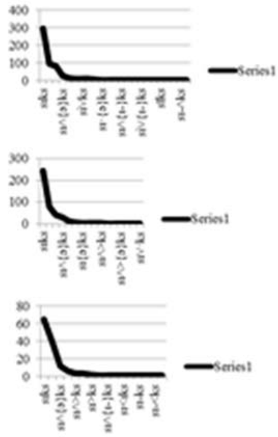
The word "law," however, is really an overstatement; the distributions in my Atlas data are always nonlinear, but quite variable in proportions, and so not subject to exact description with a precise formula, as we might have in physics for the Law of Gravity or Thermodynamics. Thus the so-called "80/20 Rule" is more 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, when the actual proportions may well be 70/30 or, as here for *thunderstorm* words, about 90/10. The A-curve is a perfect example of the nonlinear distributional pattern characteristic of complex systems.



Another hallmark of complex systems is the property of scaling, or "scale-free networks" Scaling in speech takes the form of repeating nonlinear distributions of variants for the data overall and for every subsample. Here, we see the distribution of variant realizations of the [ɪ] vowel in the word *six*, for my entire Linguistic Atlas survey in the Eastern States. Here, we see the distribution of the same data for our three "types" of speakers. Our three "Type" classifications describe levels of education and social involvement, where Type I is the lowest and Type III is the highest. Each of these charts of the [ɪ] vowel have about 80% of their tokens in the three top-ranked types.

Tallies for LAMSAS Type I, II, III

Type I (564)		sɪ^·{ə}ks	1	sɪ^·ks	4	sɪ{ə}ks	1
sɪks	295	sɪ^>·ks	1	sɪ^>ks	3	sɪ^{ɪ-}ks	1
sɪ^ks	94	sĩks	1	sɪ˅{ə}ks	2	sɪ·{ɪ-}ks	1
sɪ·ks	81	sɪ<ks	1	sɪ-<ks	2	sɪ<ɪks	1
sɪ^{ə}ks	28	sɪ-<ks	1	sɪ^>{ə}ks	1	sĩks	1
sɪ^·ks	12	sɪ-^ks	1	sɪ^{ɪ-}ks	1	sɪ-ks	1
sɪ>ks	10	sɪ-˅ks	1	sɪ·{ə>}ks	1	si˅ks	1
si˅ks	10			sɪ˅·ks	1	sɪ-<ks	1
sɪ-ks	8	Type II (442)				sɪ^<ks	1
sɪ{ə}ks	5	sɪks	246	Type III (140)			
sɪ·{ə}ks	3	sɪ^ks	79	sɪks	65		
sɪ^>ks	3	sɪ·ks	43	sɪ^ks	40		
sɪ˅{ə}ks	2	sɪ^{ə}ks	29	sɪ^{ə}ks	12		
sɪ^{ɪ-}ks	2	sɪ>ks	12	sɪ·ks	6		
sɪ˅>ks	2	sɪ-ks	8	sɪ^>ks	3		
sɪ{ə>}ks	1	sɪ{ə}ks	5	sɪ˅>ks	3		
si˅{ɪ-}ks	1	si˅ks	5	sɪ>ks	2		

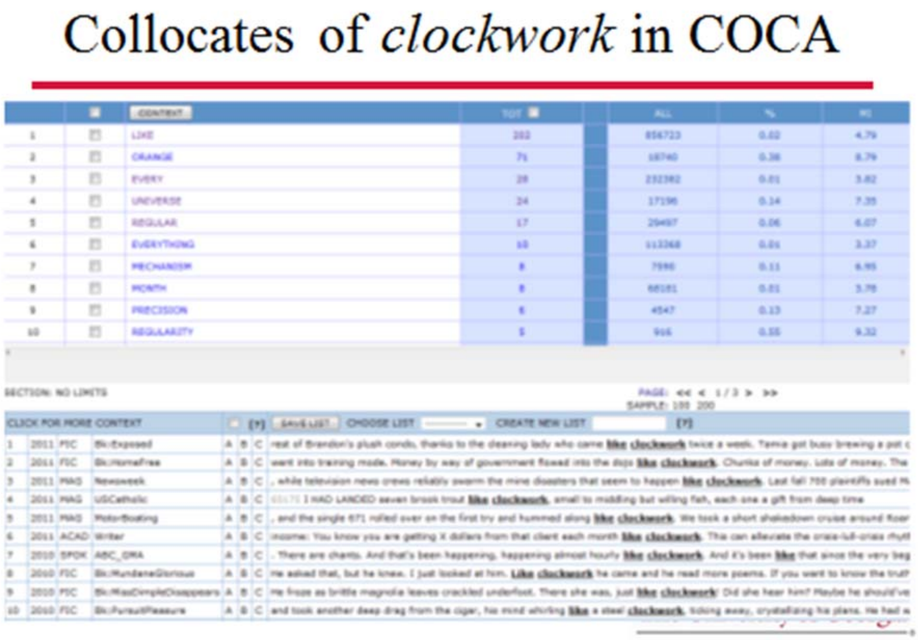


The University of Georgia

The curves have subtle differences, but the nonlinear shape appears clearly in each graph, overall and in both subsamples--and throughout all of my Atlas survey data.

So, complexity science tells us about the process by which order emerges from massive numbers of random interactions among the components in the complex system of speech, rather than from a simple cause or a set of rules or any controlling agent. As I have shown you here, whether we are talking about ants, physics, or speech, a few simple principles can give rise to frequency-based patterns. These patterns are not the same as what we usually call grammar (more on that soon). Instead, complexity science defines the relationship between language in use and any generalizations we may wish to make from it. It allows us, really for the first time, to get away from reliance on our perceptions of the language around us, and instead make effective use of the actual patterns in speech as people use it in order to resolve practical problems in language study.

To illustrate these points, let's now return to a linguistic analogue of the Corpus Clock: a brief historical analysis of the word *clockwork* in Mark Davies' American English corpora (http://corpus.byu.edu).



In his contemporary COCA corpus (425 million words from 1990-2011), the word *clockwork* occurs 468 times, and it is accompanied by these collocates. The list clearly follows the nonlinear curve we expect of a complex system, though the long tail of single occurrences is cut off is this slide. We can set aside for the moment the collocation "clockwork orange" as the creative title of a popular book and movie. The largest use of *clockwork* comes with *like*, and with words such as *every* with a time unit (as in "every month like clockwork"), or *everything* (as in "everything ran like clockwork"). Similar to such usages, we have *regular* (as in "regular as clockwork"). These are all similes, marked metaphorical uses that understand *clockwork* to mean the epitome of control. Perhaps the title "clockwork orange" plays creatively with this meaning as well. The same might be said for the unmarked metaphorical use, "clockwork universe," which understands the universe to operate by Newton's laws as regularly as a clock runs. As against these metaphorical uses, we have *mechanism*, which refers to the literal wheels and gears of a clock or of a machine like a clock. The point here is to show that our contemporary understanding of *clockwork* is not so much the literal sense, although we do have it still, but instead the metaphorical sense of regularity, whether in daily events or in the universe itself. Time passes at a constant rate, as measured more or less precisely by clockwork.

Collocates of *clockwork* in COHA

	CONTEXT	TOT	1810	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
1	REGULAR	24										3	3	2	1	4	1	2	2		5	1
2	ORANGE	13																	2	3	5	3
3	PIECE	9				1					1				1	4	1	1				
4	UNIVERSE	8														2				2	4	
5	PRECISION	7						1			1						1	2	1		1	
6	WOUND	7					1	2	1		2	1										
7	CLOCKWORK	6															6					
8	MECHANISM	5										2	2									1
9	REGULARITY	4						1			1				1						1	
10	TOY	4													1		2			1		

The University of Georgia

If we compare these results to Davies' historical COHA corpus (400 million words of American English, collected by decade over two centuries), we see that *clockwork* occurred 353 times. This must mean that its rate of occurrence has increased in recent decades. Mechanical clocks were invented in the Middle Ages, and yet the first citation of *clockwork* in the OED is from 1628, in its metaphorical sense in the phrase "this curious clocke-worke of religion." The literal sense is only documented in OED from 1662. The rise in frequency of *clockwork* in American English as shown in COHA corresponds in time to the mass production of mechanical clocks in the 19th century (it was one of the initial such industries in America). The unusual facts of the time lag in the appearance of *clockwork*, and of the appearance of the metaphorical sense before the literal sense in OED, and its relatively sudden expansion in use in American English are topics to which I will return shortly.

In the COHA list we again see the nonlinear curve we expect from complex systems in the distribution of collocates, with many of the same words. However, we do see the operation of historical change. In the 19th century, *piece, wound,* and *mechanism* are much more common than the metaphorical uses with *regularity* and *precision*. The typical meaning of *clockwork* changed during the 20th century so that the metaphorical sense became more common than the literal sense. Both of these senses were possible in each century; what changed was their frequency of use.

If we look a little closer at *clockwork universe*, however, between its emergence in COHA in 1940 and its greater prevalence in the last two decades in COCA, we see that almost all of the examples are used in a negative sense: they refer to the demise of 17th century Newtonian ideas of a regular, controlled universe in favor of the relativism of Einstein and Hawking. So, for example, people are writing things in reference to quantum mechanics like "That new theory replaced the perfectly predictable, **clockwork** **universe** of Isaac Newton with a mysterious world" (COHA 1980s token). So, in another historical development, it appears that in the most recent time period, we have come to reject the *clockwork* metaphor when it refers to our universe. In a particular domain, in usage at one small scale, we find that the meaning of the word does not follow the frequency pattern that it does more generally at the time. Again as we would expect of a complex system, we need to take scale into account in our observations, to understand that linguistic behavior may be different in different subsets of our data, in different situations of use. Furthermore, if we do see different linguistic behavior in different places or in different text types, nothing has gone wrong; such differences are not evidence of some error by the writers or of some error in our analysis. *Clockwork* not only has different frequencies of its meanings over time, but those meanings occur at different frequencies in different domains at the same moment in time. We now see that, because of the operation of speech as a complex system, and in line with our modern view of a "clockwork universe," this semantic aspect of the history of the language does not operate like clockwork. It took too long, we might think, for *clockwork* to appear in American English, and the word has had a complicated and changing semantic history. Ironically, the history of *clockwork* shows us how time appears to speed up and slow down, on occasion even to stop (in the delayed appearance of the word), when we inspect the continuous history of the word's semantic development.

Now let's consider time over the long term not for semantics but as regards the topic of grammaticalization in language change. The origin of the idea is Paul Hopper's influential 1987 paper, "Emergent Grammar." He did not derive his term "emergent" from complexity science, but his position accords exactly with complex systems as described here:

> 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' (Clifford 1986:19). 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 took the term *emergence* from cultural anthropology, from an article published about the same time the Santa Fe Institute was established to study complex systems, an example of how similar ideas can be "in the air" at a moment in time.[i] 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, much-cited phrase, this means that we can conceive of grammar as a rational object, but one that is always 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, then, 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.

This statement might be modified 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.

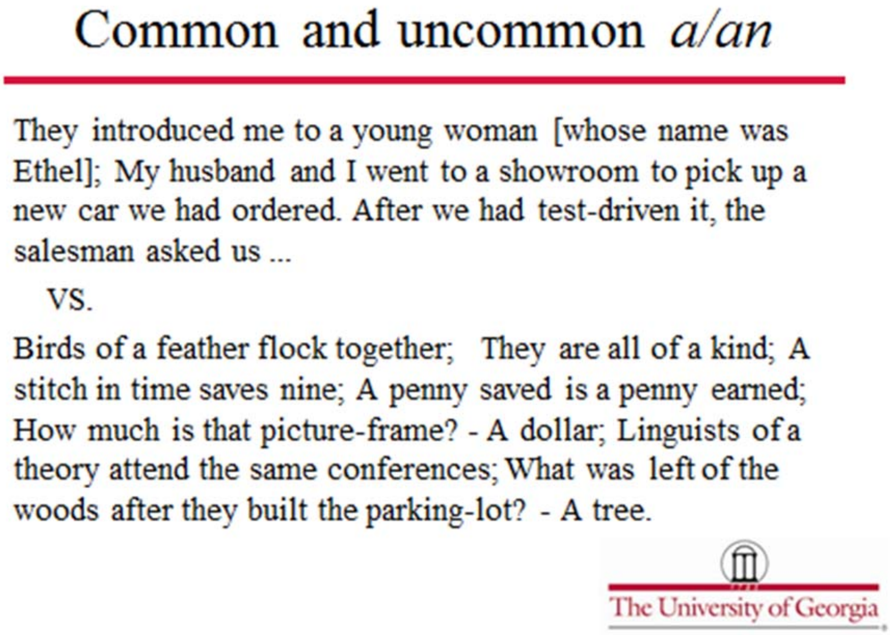
In later years, the idea of grammaticalization became much more aligned with formal linguistics, and less with Hopper's idea of emergence that connects so well with complex systems. Hopper and co-author Elizabeth Traugott (1993) define grammaticalization as a process, but now one characterized by how the properties of sentences "come into being" or have synchronic organization. Grammar has become an object, and individual constructions are "selected" as grammatical parts of it with reference to a framework, a set of categories which are non-discrete.[ii] What may have begun as emergence in a complex system has now become highly formal, reified as a structure, moreover one which is thought to participate in the continuing process. So, in *Grammaticalization*, what had started in "Emergent Grammar" the article has become radically different, and in so doing the most brilliant insight of Hopper's article, the idea of grammar as continual movement, has been returned in the book to a much more mainstream linguistic discussion.[iii]

For historians of the language, the key question is how to accommodate Hopper's sense of continual movement, and still be able to describe the grammar of the language at any moment in time. The best answer from complexity science, again, is to observe the nonlinear distribution and scaling properties of complex systems, and to put the 80/20 Rule to good practical use, and these notions lead me to suggest some improvements in our analytical practices for historical studies. The idea of selection should be reserved for rule-based generative grammars that need to "select" the most frequent forms of constructions in order to preserve the elegance of their logical systems. Generative grammars can afford to, and therefore should, ignore the 80% of possible constructions that occur rarely in order to concentrate on the 20% of the construction types that account for 80% of the tokens. To do otherwise results in "rule creep," the inelegant addition of more and more rules to account for infrequent cases, and thus also the loss of what best distinguishes generativism from structuralism. On the other hand, in structural grammars that collect paradigmatic lists of possible constructions, there is no good linguistic reason to privilege the most common variants as having been "selected" and therefore have status as being “grammatical" and to relegate less-common variants to “noise” in the system. It will be enough to note whether any particular variant is in the 80% core or the 20% periphery of constructions. All of the variants on the A-curve are actually just as relevant for inclusion in the structural system. The notion of language-internal selection, then, cannot operate within the A-curve because nothing is really chosen or preferred. For the historian of the language, this means that "change" can no longer be defined as a different variant having been "selected" at a later moment in time.

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 raised just this issue in "Emergent Grammar" using 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.

In other words, English has preserved the historical functions of *a/an*, as well as developing new ones.



As Hopper admitted, there are retentions of archaisms in proverbial language, which is where Hopper was able to illustrate modern usage of the older functions. Still, whether in proverbial language or in special domains like law or religion, we do still employ 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. 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. 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 the dangerous sense is of course the top-ranked meaning on the modern nonlinear curve of meanings for a word.

Our corpus analysis of *clockwork* showed us this process in operation in contemporary time: we can still use the word to refer to gears and wheels, and the dangerous sense of the word is regularity and precision, but we need to be aware that now we often speak of *clockwork* in a negative sense, as with *universe* and *orange*, that may give the word an unsavory or ironic semantic prosody. Thus, for grammatical constructions, 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 nonlinear distribution of contemporary usage, and that such usage continues actively to change.

For historians of the language, this means that "change" in the complex system of speech or in structural grammars will consist of an alteration in frequency of any particular feature, instead of the selection of one form over another. One way to track feature frequency, the S-curve, has already been described for the progress of linguistic change (notably in Kroch 1989, Labov 1994:65-67), and the A-curve distribution is in no way at odds with the S-curve. The two curves are actually different expressions of the same basic distributional facts. The S-curve just describes the successive frequencies of a single variant at different moments in time.

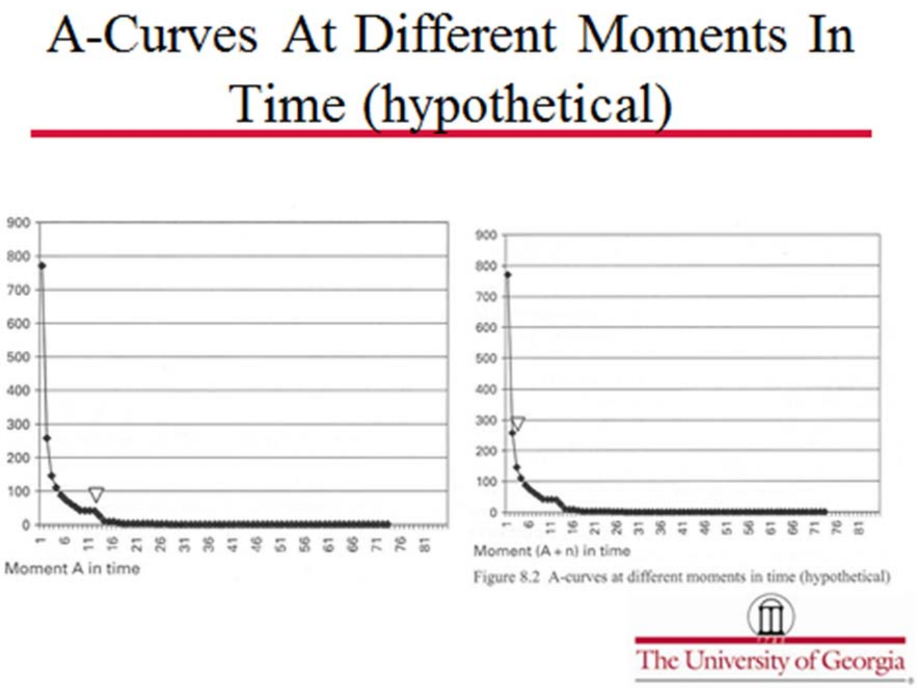
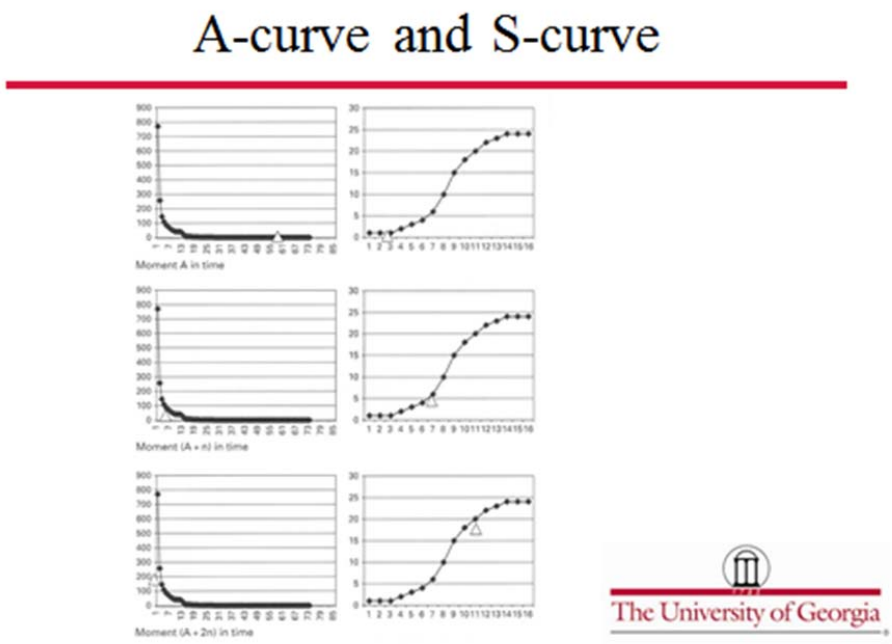


Figure 8.2 A-curves at different moments in time (hypothetical)

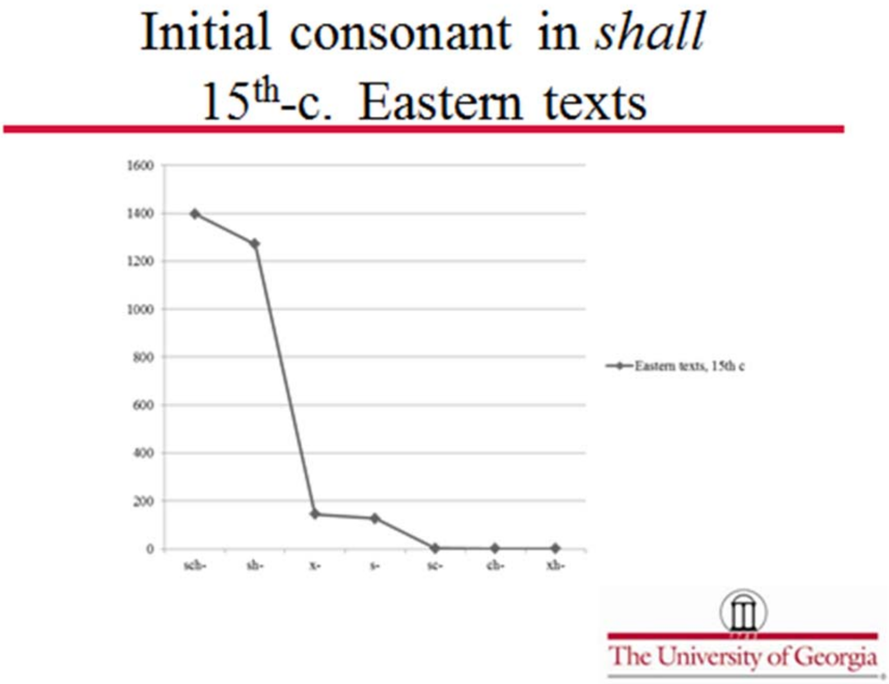
Here we see two different A-curves that correspond to different moments in time for the same variant, and locates the position of the variant on each curve. We can recall, for instance, the changing frequencies of the meanings of *clockwork* over the past two centuries.



If we then draw A-curve charts that correspond to different times on a given S-curve chart, we can see that the characteristically sudden positive movement of variants on the S-curve is mirrored by the shape of the A-curve. When a particular variant "climbs" the A-curve by moving up in frequency rank, the distributional patterns of both the A-curve and the S-curve predict that there is a larger relative change in the middle of the curves. As a variant moves from one rank to the next in the middle of the curve, each step in rank describes an increasingly large number of occurrences. This means that the small number of occurrences when a variant begins to become more frequent will be expressed as a slowly growing proportion, and the larger number of occurrences in the middle of the curve will be expressed as a rapidly growing proportion. The A-curve and the S-curve are thus complementary descriptions of the distributional facts of variant linguistic forms at different moments in time, new and improved ways to document historical change. And this focus on frequency distributions, rather than qualitative change, also allows explicitly for what Laura Wright (2000: preface) has waggishly called “W curves” that describe increases and decreases in the frequency of forms over time.

The A-curve offers historians of the language a way to judge their sparse data. Historical studies often cannot acquire enough data for a good survey, and the data that is available does not constitute anything like a valid sample of the speech or texts in use at the time of the study. Still, the 80/20 Rule offers a metric to estimate the status of forms from the limited evidence available. This means that it is highly likely that a single occurrence of a form will come from the top-ranked variants on the curve (if we could make one), but there is also a decent chance, 20%, that it is actually an uncommon use somewhere in the long tail of the curve. Similarly, if we have a small number of occurrences of variant forms, it is likely, but not certain, that the distribution will begin to show a nonlinear curve. Therefore, when the amount of evidence is small, we should especially avoid the temptation to make categorical interpretations, i.e. that the form we found necessarily was the most common one at the time, or to assume that the results are necessarily typical of the domain that we searched. The best interpretation will be one that assumes the non-linear pattern and attempts to fit the actual returns to that pattern as well as possible. Results from small data sets will not show us the whole picture, but the 80/20 Rule gives a chance to make a reasonable estimate. And we should always recognize that, because of scaling, findings from any particular domain will not yield a result that is always generalizable to larger or smaller domains.

Merja Stenroos and I have offered a good example of this kind of reasoning, with application also to aggregations of features in geographical areas for Middle English dialects (forthcoming), the first consonantal element of the verbs *shall* and *should*. In the Middle English Grammar Corpus (Stenroos, Mäkinen, Horobin and Smith, 2011), which consists of 405 text samples from the fourteenth and fifteenth centuries, eleven different spellings are found. Of these forms, only three or four occur with a reasonable frequency in any given area and period; however, the dominant forms vary greatly between areas and periods.

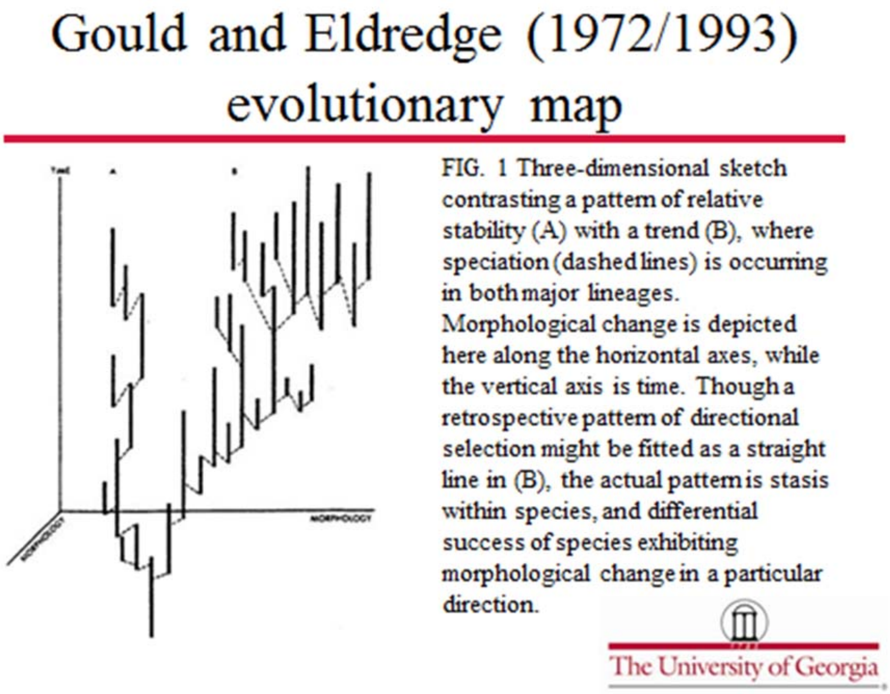


The frequency distribution in the texts localized in the Eastern part of the country and dated to the fifteenth century shows that the spelling <sch> is clearly the dominant one, while we know that <sh> eventually moves up to become the most frequent variant at the larger national scale in England. The most common spellings in the thirteenth century, <sc> and <s>, have, on the other hand, moved down the curve, surviving in the long tail of occasional forms found in any corpus of reasonable size.

Evidence like this of the operation of complex systems does not accord well with our traditional means of describing geographical varieties. Traditional dialect surveys in England and America have both posited the existence of dialects and used survey evidence selectively to show where they might be (Kretzschmar 2009: 66-74). Similarly, the “fit-technique” developed for Middle English dialects by McIntosh (McIntosh et al. 1986: 23) localizes texts based on the assumption that Middle English linguistic variation forms a regular continuum, into which any dialectally consistent text may be placed. Both for Middle English and for modern English, these traditional approaches apply a formal assumption of the regularity of a dialect to evidence whose variation is anything but regular. On the other hand, when we apply our knowledge of complex systems, Merja and I would prefer to map texts according to their provenance, as far as retrievable, arriving at a messier picture that simply answers the question “who wrote what where.” The study of texts associated with a particular location allows us to reconstruct something of the sociolinguistic reality within the community, which may have been complex (see Stenroos and Thengs 2011). At the same time, while some scribes travelled and some reproduced the forms of non-local exemplars, we expect that the local variants will turn out to be most frequent overall in the A-curve of variant frequencies--like the one for the *shall* spellings in Eastern texts. Thus, the "fit-technique" in Middle English dialectology, like traditional modern dialectology, is best thought of as a method dominated by its formal assumptions. This does not make it bad in itself, but it does mean that, as was the case for the construction of grammars, we need to understand the degree of abstraction of the "fit-technique" as it defines one dialect, one grammar, to fit all the texts of a place--because complex systems tells us that that is never actually the case in the evidence itself.

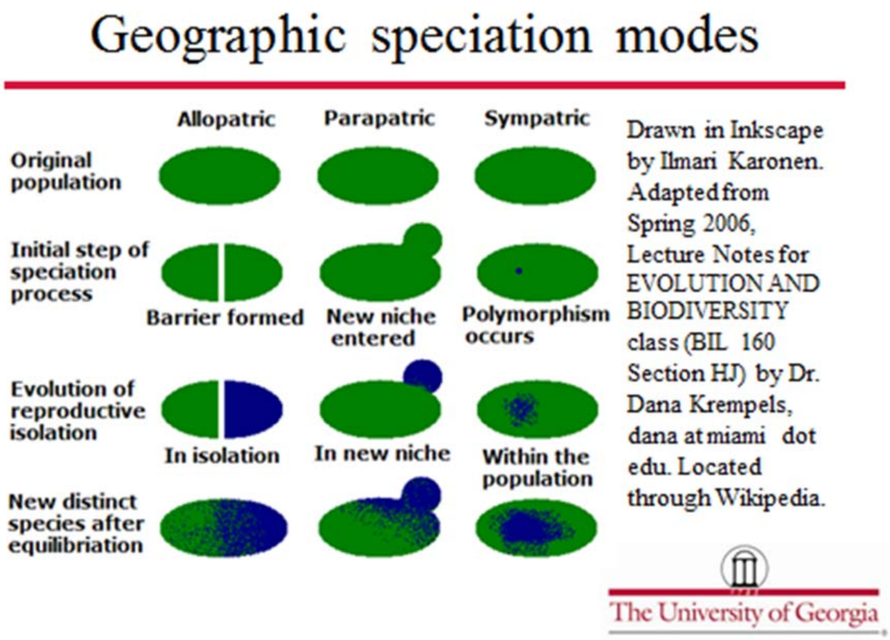
For my final point, I would like to pick up the geographical thread from our discussion of Middle English dialectology, and combine it with what I promised to say about the Corpus Clock. The clock, I said, is only accurate every five minutes: time slows down for a while, and then speeds to catch up. It turns out that this seemingly perverse aspect of the clock is actually a good model for what we do with long-term change in historical linguistics, because it matches a model for biological evolution, called "punctuated equilibrium," that capitalizes on how complex systems work. In linguistics, many others have discussed potential parallels between biological evolution and language change, most often to reject them (e.g. Labov 2001, Lass 1990). Salikoko Mufwene has offered the latest influential parallel (2001, 2008), in which he makes the analogy of language to a "parasitic species." I would like here to say, not that evolutionary biology in the form of punctuated equilibrium provides an apt metaphor for language change, but that the workings of a complex system actually account for exactly the same process in both biology and language over time.

Gould and Eldredge proposed the idea of punctuated equilibrium in 1972. As they reported in a later review article (1993), their theory has now received wide acceptance after an initial period of misunderstanding.[iv] They proposed that evolution was not a slow, gradual, process that operated like clockwork, but instead that species enjoyed long periods of stasis during which sub-populations of the species came to coexist, whether through random genetic drift or because of differences in environmental conditions. Our modern view of this process is affected by what they call "the notorious imperfection of the fossil record" (1993: 222), that is, the fact that we find fossils at widely separated locations representing significantly different times, rather than collecting fossils that represent a valid sample over time of the central population of the species and of any sub-populations that may have developed. As Gould and Eldredge put it in paleontological terms, "small populations speciating away from a central mass in tens or hundreds of thousands of years, will translate in almost every geological circumstance as a punctuation on a bedding plane, not gradual change up a hill of sediment, whereas stasis should characterize the long and recoverable history of successful central populations" (1993: 222). To be clear, Gould and Eldredge are pointing out that different species will be present in the same stratum of fossils, because of the long static periods that each species can exist and because of coexistent sub-populations of the species. Before punctuated equilibrium, it was thought that up to 90% of the differences in fossils came from the gradual evolutionary process but now, they report, "all substantial evolutionary change must be reconceived as higher-level sorting based on differential success of certain kinds of stable species, rather than as progressive transformation within lineages" (Gould and Eldredge 1993: 223). And, to get back to the Corpus Clock, the accidents of what fossils we find can make it look like time speeds up and slows down, that change can appear to happen quickly sometimes and at other times can seem to stop. Punctuated equilibrium looks like syncopated time.



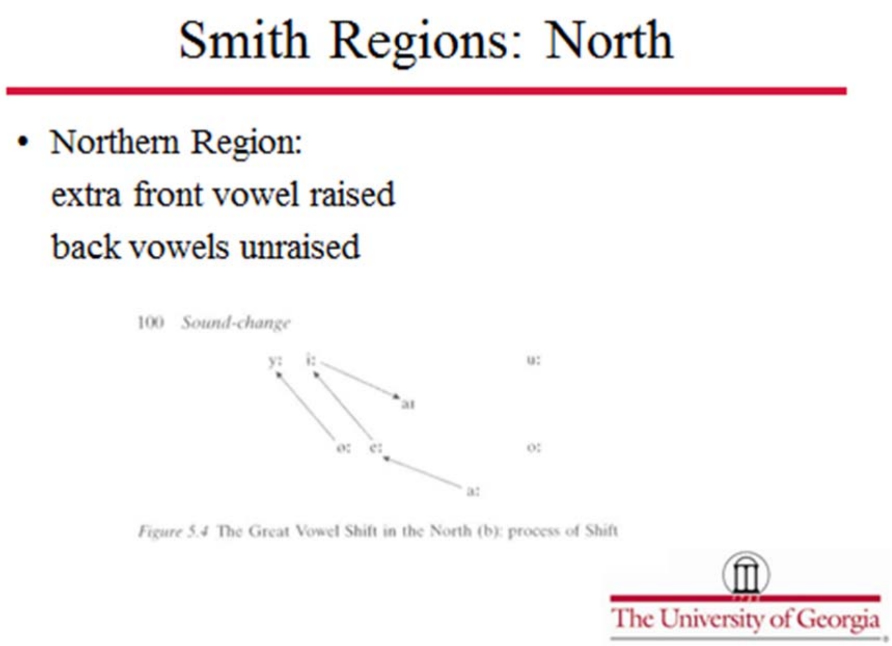
The Gould and Eldredge map of the evolutionary process contrasts the stability of the population at left (which represents gradual phylogenetic change), with the greater speciation of populations at right (which represents the simultaneous development of sub-populations). As I just suggested, the appearance of sudden jumps in evolution comes about because imperfections in the geological fossil record may lead us to miscategorize the simultaneous existence of different populations as sequential populations. If we were to draw a straight line across the graph from any point on the vertical Time access (actually if we draw a plane, since this is meant to be a 3D chart), it would strike multiple populations from the sub-populations at right. Thus, from observations at any given moment in time, we are nearly certain to retrieve evidence that belongs to different populations, and the trick for us is to make an appropriate analysis of that evidence. That is, we need to deal with the appearance of slow change and fast change that emerges from such an array of facts. What must be clear, however, is that the evidence we collect from any fossil record that corresponds to the right side of the chart cannot come from a single, continuous, gradual phylogenetic process. To try to make the right side of the chart look like the left side would make us discount some of the evidence we have, or prefer some parts of the evidence over other parts, in order to force the evidence to make the neat generalization of a gradual, clockwork evolutionary process.

What Gould and Eldredge could not know in 1972, but what Gould had realized by the time he wrote his last book (2003), is that the right side of the chart corresponds exactly to what we would expect of a complex system. The scale-free property of complex systems predicts that there will always be sub-populations at different scales. We have already seen this in the distribution of individual features from the linguistic data; now we are just observing the same property at a higher level of scale of species, or in linguistic terms, in the aggregations of features we might call a dialect or a language.

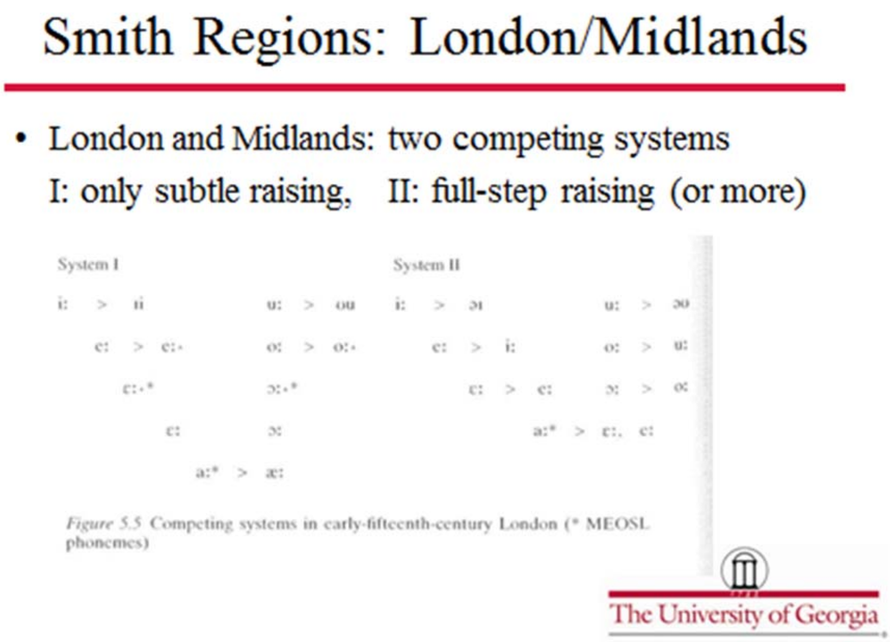


Gould and Eldredge originally matched their sub-populations to different geographic populations, known as allopatric, parapatric, and sympatric modes of speciation. In each of these cases, a sub-population occurs in a geographic niche, whether created by a barrier, an extension, or an internal division. What we now understand from complex systems is that the pattern of sympatric speciation is not limited to geographic boundaries, although these remain an important consideration, but that speciation can occur for other reasons as well: the scale-free property of the complex system predicts that there will always be many sub-populations, each one defined by its own nonlinear distribution of morphological characteristics. Complex systems also tells us that it will always also be possible to consider populations at higher levels of scale, to consider any combination of sub-populations we want right up to the population as a whole, and that at every level of scale the different morphology of the sub-populations will be expressed in the nonlinear distribution of characteristics. We see this represented at the bottom of the chart, "after equilibration," when individuals from a particular sub-population are mixed in with the population as a whole. Thus, classic examples of geographical speciation like Darwin observed in the finches of the Galapagos do involve geography, but our new knowledge of complex systems tells us that geography is not the whole story.

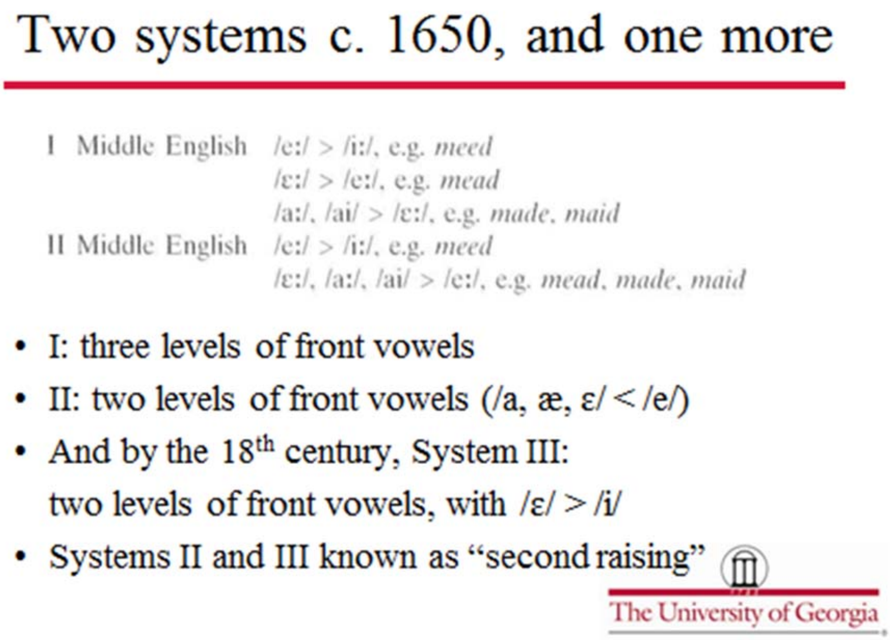
An excellent example of the process of punctuated equilibrium in the History of English is the Great Vowel Shift, and Jeremy Smith has shown it to us (1996).[v] The general outline of the Great Vowel Shift should be clear to this audience, including the raising of both the long front and long back vowels.



Smith, however, points out that in the North, the Great Vowel Shift did not proceed according to the usual generalization: an extra front vowel was raised, and the back vowels remained unraised. Smith offers us complications of the Great Vowel Shift in the South, too.



He describes two competing systems in 15th-century London, the first among those whom he calls "descendants of Chaucer," from higher social circles, and the second system prevailing among the remaining speakers in the Midlands and South. We see here geographical sub-populations, but also a pair of social sub-populations within a region.



Somewhat later in his argument, Smith also introduces a third system which originated in East Anglia but whose speakers also went to London, where their speech pattern competed with that of the descendants of Chaucer; speakers retained System II longer in the countryside. This amounts to a kind of linguistic equilibration, as described earlier for modes of speciation. I recommend the orthoepic evidence and details of Smith's argument to you from his book. My point here, of course, is that complex systems predicts that both the regional and social sub-populations could develop different characteristics, and that these sub-populations could coexist at the same time, and that, just as in punctuated equilibrium, we can observe "equilibration" as members from one sub-population become mixed with others. So, the linguistic evidence is not just parallel to the fossil record. It is the same procedure to use the fossil record to describe species as it is to use fragmentary linguistic evidence to describe competing dialects in English, or in the longest historical perspective, competing languages in the Indo-

European family. The same underlying process is at work for both paleontology and historical linguistics. To rephrase Gould and Eldredge, while we used to worry about dialect contamination in the gradual evolution of a language, now all substantial evolutionary linguistic change ought to be reconceived as higher-level sorting based on differential success of certain kinds of stable language varieties.

Given Smith's evidence and my argument, then, what should we make of the Great Vowel Shift? The complex systems approach would suggest that we should not abandon it: the Great Vowel Shift remains a useful generalization at the top level of scale. But we cannot just work at the top level of scale, because we are then subject to the dangers of misinterpretation of our imperfect evidence that come from punctuated equilibrium. Just as for the fit-technique, we cannot apply the top-level generalization of the Great Vowel Shift back down onto primary evidence, because it is based on the formal abstraction that we can define one grammar to fit all the texts of a place--and complex systems tells us that that is never actually the case.

Just as we must do with the Cambridge Corpus Clock, we have to expect that time in some places will appear to go faster and in some places slower in historical linguistics. We now know that linguistic change does not just run like clockwork. There were good reasons why the Neogrammarians created a mechanical process for linguistic change, chiefly to adopt the best science then available to govern their work. We, too, need to do this: we should not just accept traditional methods because they are traditional; we should adopt the best science available now to govern our work. I am pleased to offer you complex systems for that purpose. If we choose to make formal generalizations, and in historical linguistics a great many of us do, then any formal statements we make should reflect what we now know about the emergence in our evidence of scaling, nonlinear frequency patterns from the complex system of speech.

References

Beckner, C., Blythe, R., Bybee, J., Christiansen, M., Croft, W., Ellis, N., et al. 2009. Language is a Complex Adaptive System: Position paper. *Language Learning* 59, Supplement: 1-26.

Eldredge, Niles, and Stephen J. Gould. 1972. Punctuated Equilibria: An Alternative to Phyletic Gradualism. In *Models in Paleobiology,* ed. T. Schopf (San Francisco: Freeman, Cooper), 82-115.

Gardner, Martin. 1970. Mathematical Games: The fantastic combinations of John Conway's new solitaire game "Life." *Scientific American* **223**: 120–123.

Gould, Stephen Jay. 2003. *The Hedgehog, the Fox, and the Magister's Pox: Mending the Gap Between Science and the Humanities.* New York: Three Rivers Press.

Gould, Stephen J., and Niles Eldredge. 1993. Punctuated Equilibrium Comes of Age. *Nature* 366 (18 Nov): 222-227.

Hawking, Stephen. 1988. *A Brief History of Time.* London: Bantam.

Hawking, Stephen, and Leonard Mlodinow. 2010. *The Grand Design.* New York: Bantam.

Holland, John. 1998. *Emergence: From Chaos to Order*. New York: Basic.

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. 2009. *The Linguistics of Speech*. Cambridge: Cambridge University Press.

Kretzschmar, William A., Jr., and Merja Stenroos. Forthcoming. Evidence from Surveys and Atlases in the History of the English Language. To appear in Terttu Nevalainen and Elizabeth Traugott, eds., *The Oxford Handbook of the History of English* (Oxford: Oxford University Press).

Kroch, Anthony. 1989. Reflexes of Grammar in Patterns of Language Change. *Language Variation and Change* 1: 199-244.

Labov, William. 1994. *Principles of Linguistic Change: Internal Factors*. Language in Society, 20. Oxford: Basil Blackwell.

Labov, William. 2001. *Principles of Linguistic Change: Socal Factors*. Language in Society, 29. Oxford: Basil Blackwell.

Lass, Roger. 1990. How to Do Things with Junk: Exaptation in Language Evolution. *Journal of Linguistics* 26: 79-102.

Lewis, C. S. 1990. *Studies in Words.* 2nd ed. Cambridge: Cambridge University Press.

McIntosh, Angus, Michael L. Samuels and Michael Benskin. 1986. *A Linguistic Atlas of Late Medieval English,* Aberdeen: University Press.

Mitchell, Melanie. 2009. *Complexity: A Guided Tour.* Oxford: Oxford University Press.

Mufwene, Salikoko. 2001. *The Ecology of Language Evolution.* Cambridge: Cambridge University Press.

Mufwene, Salikoko. 2008. *Language Evolution: Contact, Competition, and Change.* London: Continuum Press.

Schneider, Edgar. 2007. *Postcolonial English: Varieties around the World.* Cambridge: Cambridge University Press.
Séguy, J. 1971. La relation entre la distance spatiale et la distance lexicale', *Revue de Linguistique Romane* 35: 335-357.
Smith, Jeremy. 1996. *An Historical Study of English*. London: Routledge.
Stenroos, Merja, Martti Mäkinen, Simon Horobin and Jeremy J. Smith (compilers), 2011. *The Middle English Grammar Corpus*, version 2011.1. January 2011, University of Stavanger, http://www.uis.no/research/culture/the_middle_english_grammar_project/meg-c/
Stenroos, Merja, and Kjetil Thengs. 2011. Two Staffordshires: Real and Typological Space in the Study of Late Middle English Linguistic Variation. Paper presented at Helsinki Corpus Festival, Helsinki.
Wright, Laura, ed. 2000. *The Development of Standard English 1300-1800*. Cambridge: Cambridge University Press.
Zipf, George. 1949. *Human Behavior and the Principle of Least Effort*. Cambridge, MA: Addison Wesley.

Abstract

Complexity theory (Mitchell 2009, Kretzschmar 2009) is something that historical linguists not only *can* use but *should* use in order to improve the relationship between the speech we observe in historical settings and the generalizations we make from it. Complex systems, as described in physics, ecology, and many other sciences, are made up of massive numbers of components interacting with one another, and this results in self-organization and emergent order. 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 spatial and social and textual groups in which we actually communicate. Order emerges from such systems by means of self-organization, but the order that arises from speech is not the same as what linguists study under the rubric of linguistic structure. In both texts and regional/social groups, the frequency distribution of features occurs as the same pattern: an asymptotic hyperbolic curve (or "A-curve"). Formal linguistic systems, grammars, are thus not the direct result of the complex system, and historical linguists must use complexity to mediate between the language production observed in the community and the grammars we describe.

As for applications to historical linguistics, first, the scaling property of complex systems tells us that there are no representative speakers, and so our observation of any small group of speakers is unlikely to represent any group at a larger scale—and limited evidence is the necessary condition of many of our historical studies. The fact that underlying complex distributions follow the 80/20 rule, i.e. 80% of the word tokens in a data set will be instances of only 20% of the word types, while the other 80% of the word types will amount to only 20% of the tokens, gives us an effective tool for estimating the status of historical states of the language. Such a frequency-based technique is opposed to the typological "fit" technique that relies on a few texts that can be reliably located in space, and which may not account for the crosscutting effects of text type, another dimension in which the 80/20 rule applies. Besides issues of sampling, the frequency-based approach also affects how we can think about change. The A-curve immediately translates to the S-curve now used to describe linguistic change, and explains that "change" cannot reasonably be considered to be a qualitative shift. The GVS, for example, is a useful generalization, but complex systems explains why we should not expect it ever to be "complete" or to appear in the same form in different places. Finally, complexity science helps us to see and understand how English continues to "emerge" around us in the ongoing complex system of our speech, so that any process of "standardization" does not just lead inevitably to Modern English, but must be understood as a limited and highly specialized part of the history of English. These applications of complexity can help us to understand and interpret our existing studies better, and suggest how new studies can be made more valid and reliable.

[i] One of the best current treatments of complex systems and language has a central concern with this topic. The Five Graces Group, which includes Joan Bybee, Nick Ellis, John Holland, and Diane Larsen-Freeman, among others, has produced a position paper called "Language Is a Complex Adaptive System" (Beckner et al. 2009). [redacted]They begin with a basic tenet of grammaticalization (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” (Bybee, 2002). For example, the English contracted forms (I’m, they’ll) originate from the fusion of co-occurring forms (Krug, 1998). Auxiliaries become bound to their more frequent collocate, 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, is a universal process, one which describes the operation of the complex adaptive system of speech. It is taken to operate at the level of single languages. Since the Five Graces propose that grammar is "a network built up from the categorized instances of language use," then grammaticalization is the mechanism by which they propose that such a network arises and changes.

[ii] 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. The role of linguists is to select grammatical constructions in an attempt to make useful idealizations from speech data.

[iii] The Five Graces also describe grammar as a kind of object, a "network," that converts Hopper's process into a state, and identifies the state of the complex system as its grammar. 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). For any single linguistic feature, which is not so simple as a traffic light, 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.

Again, the preference for, or "fixation," of a state takes the dynamic movement of the complex system and freezes it, so that one variant of a features 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 both the benefit of understanding speech as a complex system. Of course there is nothing at all wrong with making grammars. This kind of formal analysis has been highly productive. But there is an essential conflict between making a grammar as a static hierarchy for a language at one moment in time, and the process of change in language that is best described as frequency change within a complex system.

[iv] They did not propose, as some claimed, that species suddenly leaped to new genetic configurations (the "saltational theory"), and they opposed attempts to coopt their theory by creationists.

[v] Smith was already interested in chaos theory and complexity in 1996, but did not yet have the means to describe his evidence in those terms. Of course, now we can understand from complex systems that there will be an A-curve distribution of variants for every vowel following the 80/20 Rule, and that we can assemble the individual vowel frequency patterns to make a generalization at a larger level of scale for a regional dialect.