Aspects of a Theory of Grammatical Change

Peter W. Culicover
The Ohio State University

Abstract

The question as originally posed for this workshop is “How do opposing trends towards simplicity and communicability generate language change?” To address such a question productively it is important to be as clear as possible about what actually is changing, and what the conditions are that lead to change. So I will situate my remarks in a broader theoretical context.

I briefly review the Simpler Syntax (SS) perspective on grammar and contrast it with mainstream generative grammar (MGG) (§1). I illustrate the difference with some examples of constructions and of what I have called ‘syntactic nuts.’ (§2) It is possible to model the contribution of grammatical complexity to language change with very idealized computational simulations (§3). The key factors are those that have been used to model the dynamics of social attitudes in networks of agents under various parameters settings. But such modeling ignores the foundations of grammatical complexity. This is not a simple notion – it has to do not only with the formal complexity of linguistic expressions but with the transparency of their correspondence with meaning.

With this as background, we come to the central point, which is that grammatical complexity is a central engine of grammatical change (§4). I will suggest some plausible measures of complexity with SS as background.

The two forces of complexity and transparency, with attendant mechanisms, lead to a view of change as cyclic, where some constructions are generalizing over time towards rules, while some rules are contracting over time.

---

1 This paper was prepared for the Santa Fe Workshop on Building Integrated Models of Linguistic Change. Thanks go to the organizers, Morton Christiansen, Dan Hruschka and Steve Lansing for inviting me and giving me the opportunity to organize this material for the workshop. I am extremely in debt to my collaborators Ray Jackendoff, Andrzej Nowak and Wojtek Borkowski for making it possible for me to talk about these matters – their influence and contributions to our joint work can be seen throughout, although I take full responsibility for what I have done with it here. Finally, I want to thank Erhard Hinrichs and Marga Reis and the Alexander von Humboldt Stiftung for providing me with the time to develop my thoughts.
towards constructions and even syntactic nuts. I illustrate this point with some observations about the history of English *do*-support and related phenomena (§5).

1. Simpler Syntax and Mainstream Generative Grammar

I begin with the uncontroversial assumption that the function of a grammar of a natural language is to account for the relationship between sound and meaning. All formal theories distinguish phonology, which concerns the organization of sound, semantics and pragmatics, which have to do with literal and contextually determined meaning, and syntax, which is the interface between sound and meaning. The lexicon is the repository of unpredictable sound/meaning correspondences, and morphology concerns the structure of words and how their structure bears on their meaning and their syntactic functions.

In the first section of *Simpler Syntax*, Ray Jackendoff and I undertook a survey of the historical and conceptual development of mainstream generative grammar (MGG). Our concern was to understand where there were clear empirical motivations for innovation in the development of the successive versions of mainstream theory, and to what extent various aspects of these theories were motivated instead on the basis of general methodological principles that were perhaps not even well articulated. To make a long story short, we argue that much of the development of MGG is motivated by the quest for uniformity, in many if not most cases independently of any demanding empirical considerations.

This development is understandable if we recognize that the syntactic component was the most clearly articulated part of the sound/meaning mapping available at the outset of generative grammar. It was therefore natural to syntactically as much as possible, and to insist on simplicity and uniformity of syntactic formulations as a way of capturing generalizations about this mapping. We call this tendency ‘syntactocentricity’. Moreover, because of syntactic uniformity, particular syntactic structures are posited that are not necessarily independently motivated by the facts. This is an important observation when we are concerned with change, because the architecture of the grammar that we assume tends to drive the kinds of scenarios that we develop when we are looking to explain language change.

To take a simple example, in sentence (1), an example of ‘control’, one understands Ozzie to be not only the ‘tryer’ but also the ‘drinker’, even though the noun phrase Ozzie is not overtly an argument of the verb *drink*.
Ozzie tried not to drink.

The masterstroke behind mainstream generative grammar was to propose that the missing piece of meaning is supplied by an element in a covert level of syntactic structure (‘Deep Structure’ in early work, later ‘Logical Form’). Sentence (1) has the covert form (2), in which the verb drink actually does have a subject – PRO, an unpronounced pronoun whose antecedent is Ozzie.

(1) Ozzie tried not to drink.

(2) Ozzie tried [PRO not to drink].

This analysis follows what we refer to as Interface Uniformity, which assumes that the same meaning corresponds to the same syntactic configuration.

Such an approach is effective – and appealing – for relatively straightforward situations such as (1). However, we show that carrying this strategy through systematically leads to unwelcome consequences, and that the most satisfactory account of control must be stated in terms of relations between the semantic arguments of the relations that correspond to the verbs. (There may, however, be syntactic constraints on the control relation.)

More generally, Simpler Syntax suggests that it may be profitable to adopt an alternative perspective on the architecture of the grammar, one that has been quite robust over the years but out of the mainstream. It is that the explanatory burden is not the sole responsibility of syntax, but resides in the interfaces and correspondences between syntax, morphology, semantics, pragmatics, and phonology. Simpler Syntax proposes that the syntax should be responsible only for those aspects of language that cannot be adequately accounted for otherwise, on the grounds that morphology, semantics, pragmatics and phonology are independently motivated (if not always completely well understood). This is the Simpler Syntax Hypothesis.

Simpler Syntax Hypothesis (SSH). The most explanatory syntactic theory is one that imputes the minimum syntactic structure necessary to mediate between phonology and meaning.

By defining ‘minimum syntactic structure’ we are able to anchor the syntactic part of the account and thereby reduce significantly the space of possible analyses and thus possible explanations.
The main consequence of the SSH for the study of grammatical change is that it places very severe constraints on the syntactic analysis of any given grammatical construction. Rather than allowing a syntactic analysis that is at least as rich as the corresponding semantic interpretation (because of a systematic application of Interface Uniformity), it puts the burden of the structure-meaning correspondence on the rules of interpretation, rather than on the syntactic structure itself. The syntactic structure is the simplest structure that is sufficient to account for the meaning, given that the rules of interpretation are independently required.

The potential consequences for a theory of grammatical change are attractive. SS reduces significantly the number of alternative syntactic accounts that must be entertained, thereby narrowing the space of possible accounts of an observed change. I give an example of how this works in the next section.

2. Constructions and syntactic nuts

We argue in SS that the right level of description is the set of correspondences (3).

(3) \(<\text{PHON,SYN,CS}>\)

These can be a specific as a single word, or, in the limit, as general as a rule of compositional interpretation for a head-complement structure such as

(4) \([\text{XP} \ldots \text{X}^0 \ldots ]\).

The case has to be made for the robustness of the intermediate region, which consists of correspondences that involve some degree of syntactic and semantic complexity that cannot be reduced to the correspondences of individual words or those of fully general structures.

Such intermediate correspondences are usually referred to as ‘idioms’ or ‘constructions’ depending on how general they are. SS discusses a number of cases, *Syntactic Nuts* (SN) (Culicover 1999) a number of others, and there are additional cases discussed in the literature under the general rubric Construction Grammar. The key characteristic of a construction that distinguishes it from a rule, in our way of thinking, is that the meaning of a construction is not an instance of Fregean Compositionalty (FC).
FC: ‘The meaning of a compound expression is a function of the meaning of its parts and of the syntactic rules by which they are combined’ [Partee et al. 1990]

Idioms and proverbs, such as

(5) a. The devil is in the details.
b. It takes two to tango.
c. My car kicked the bucket.

are instances where the meaning is associated with the entire expression and not to any of its parts (although technically it is possible to locate the meaning in one of the parts, of course). These are not particularly controversial in the big picture, although accounting for them is not entirely straightforward.

Another way in which FC can be violated is if there are aspects of the interpretation of an expression that are determined through the relationship between that expression and some other expression, or a state of affairs that can be interpreted. SN and SS discuss fragment constructions such as sluice stranding (also called ‘slifting’) in English, illustrated by (6).

(6) a. Otto was dancing with someone, but I don’t know who with.
b. Otto was dancing, but I don’t know who with.

The expression who with means who Otto was dancing with, but only in virtue of Otto was dancing in the preceding clause. Classical accounts contend that who with is the visible part of a complete sentential structure (7) –

(7) who Otto was dancing with

– where the material that is struck out is syntactically (and therefore semantically) real but phonetically invisible and licensed by the identical material in the antecedent. On the other hand, SS proposes that the syntactic structure of who with is the maximally simple

(8) [s who with ]

and that the interpretation Otto was dancing is supplied by a rule that draws upon the form and the meaning of the antecedent. The essence of the argument is that
such rules are needed even when the antecedent lacks the form and meaning that is required to interpret the fragment.

Another way in which FC is violated is where part of the meaning is associated an entire structural configuration and not with the parts. Some examples are given in (9).

\[
\begin{array}{l}
\text{shouldered} \\
\text{elbow} \\
\text{hammered} \\
\text{weaseled} \\
\text{danced}
\end{array}
\]

(9) Susan \[ \{ \text{shouldered} \\
\text{elbow} \\
\text{hammered} \\
\text{weaseled} \\
\text{danced} \} \] her way into a position of authority.

It seems reasonable to say that the verbs \textit{shoulder, elbow, hammer, weasel} and \textit{dance} do not inherently denote actions that involve movement along a path, and that the path interpretation is associated with the entire VP construction (Jackendoff 1992, 26). (One might even question whether they are all verbs independently of this construction.)

In all of these cases, which we consider to be constructions in the technical sense, there is a notable property, which is that the construction resembles the regular structures of the language. The idioms and proverbs look just like regular sentences and VPs, and sluice-stranding has the characteristic of preposition stranding, which is a special property of English. And languages that lack preposition stranding lack sluice-stranding, although they have regular sluicing, of the form.

(10) Otto was dancing with someone, but I don’t know with whom.

And, finally, the construction illustrated in (9) appears to be a special case of the general English VP structure, captured by the rule (11).

(11) \[ \text{VP} \rightarrow \text{V} (\text{NP}) (\text{PP}) \]

This last observation suggests something rather important about the relationship between specific constructions and the general structures of a language, which I will formulate as a general principle.
Naturalness Principle: The ‘naturalness’ of a construction in a language is determined by the extent to which it conforms to the general structure of the language.

To the extent that we can then motivate a systematic relationship between the naturalness of a construction and language acquisition (as Chomsky 1965 originally suggested in his discussion of the evaluation metric), and a relationship between language acquisition and language change, we may be able to find in the theory of constructions the basis for an account of why languages change, and why we find the kind of variation that exists. I come back to this idea in the next two sections of this paper.

The very close congruence between constructions and regular structures has led those working within the framework of uniformity to try to find ways to eliminate the apparent exceptions to FC. This can be done, in general, by assuming invisible elements that have the interpretation that cannot be located in the visible elements. One example that we have already seen is sluice-stranding. If we take a constructional approach, the meaning is in the structure itself and the interpretation rules.

SS argues that having some meaning located in the structure and the interpretation rules is independently motivated, so these are not just two equivalent ways of notating things. But even if they were, there is a sense in which the construction is the right level of description to account for the idiosyncrasies. The reason is that if the structures and derivations are uniform, the exceptionality or idiosyncrasy cannot reside in what the structure is, it must reside in the application or non-application of a rule that transforms the structure so that it acquires the proper form for the construction. Consider once again the case of sluice-stranding. It turns out that in English not all combinations of wh-phrase and preposition are well-formed instances of sluice-stranding, an observation that led Culicover 1999 to argue for its essential constructional character. E.g.,

(12) a. … but I don’t know who with.
    b. … *but I don’t know which man with.
    c. … ?but I don’t know who for.

Suppose that this construction is described as a set of possible forms with a rule of interpretation. Among these forms are

(13) who with
what with
what about
what for

The rule of interpretation ‘reconstructs’ the wh-phrase as an argument of the preposition and then it falls together with the general rule of interpretation for sluicing. It is easy to imagine the extension of the list in (13) to forms such as

(14) who for
    who about

to through extension of the pattern

(15) NP[wh, pro] P

resulting in a very restricted case of language change. Such extension is something that we would not be surprised to find, given the behavior of children when faced with exceptionality. Extension to all wh-phrases and prepositions might be possible in principle, perhaps subject to phonological/prosodic constraints, as some have suggested.

    Consider how to formulate this change in terms of a uniform treatment of sluice-stranding. I will not work through all of the details. The key is that there must be a notation on who and with that allows for the deletion of the intervening material –

(16) who[+Sluice-strand] Otto was dancing with[+Sluice-strand] ⇒
    who[+Sluice-strand] Otto was dancing with[+Sluice-strand]

-- and then extension of the feature to for, so that we get who for. But this does not take into account that both the wh-phrase and the preposition are participating in the construction, and that the extension is not to for, since for is already possible with what for. The extension is to the construct who for.

    So what I want to suggest is that explicit mention of the construction is centrally relevant to the account of acquisition and change. While it is possible to talk about acquisition and change in terms of the components of the structure and features marked on the structure, when what we are dealing with is a construction these will afford at best descriptive and not explanatory accounts of what is going on.
3. Change and variation

To go from description to explanation, we have to return to naturalness. While it is ultimately necessary to give a principled account of what is natural and general what is not, I think that it is possible to explore productively the consequences of building naturalness directly into theories of language acquisition and change while leaving the details to a promissory note. I will try to say something about the details in the next section, but what I say will necessarily be sketchy.

3.1. Introduction

Let me review very briefly the model of change developed by Culicover and Nowak 2003. In this model, we assume that language change occurs in part as the consequence of different learners being exposed to potentially different evidence regarding the precise grammar of the language that they are to learn. We assume, following Chomsky 1965, that learners chose the least complex grammar consistent with their experience, and may even overlook counterevidence to the most economical solution unless the counterevidence is particularly robust. To the extent that learners seek minimal complexity we expect language change to reflect this preference in the relative ubiquity of certain grammatical devices compared with others, and even in the appearance of universals (Briscoe 2000).

3.2. The simulation model

Here is the simulation model. There is a society of speakers of a language, some of them competent speakers and some of them learners. Each speaker interacts with some of the other speakers with some frequency, in part as a function of the distances between them. (Distance may be understood as physical and/or social.) As a consequence of drift, noise in the information channels, spontaneous change, conscious innovation and contact with other languages there will be linguistic diversity in this society. Some learners may have considerable experience with diversity, others may have very little. Over the course of generations, learners interact with speakers whose language is determined by interactions with similar speakers, so that there is a consistency of grammar that may distinguish the social group from another, more distant group.

In order to test the general properties of the interaction between language learning and language change we developed a simulation model of social interaction based on the theory of social impact due to Latané and computational
simulations based on this theory developed at the Center for Complex Systems at the Institute for Social Studies of the University of Warsaw by Andrzej Nowak and his colleagues. Our intuition was that the transmission and clustering of linguistic properties through social contact should display the essential properties of the transmission and clustering of any cognitive features. The main differences are that linguistic properties, unlike social attitudes, are not subject to conscious evaluation, and the formal properties of linguistic expressions correspond to some extent to their meanings.

3.2.1. The network of agents

We suppose for the sake of the simulation that the class of possible grammars of natural languages can be characterized entirely in terms of values of features. A prevalent view in current linguistic theory is that most if not all of the most theoretically interesting aspects of language variation, language change and language acquisition can be accounted for in terms of a small set of binary features, called ‘parameters’. For our purposes, however, it is sufficient to assume that whatever the features are, however many there are, and whatever values they have, learners are influenced to adopt the values of their community through social interaction and set their values for these features on the basis of such interaction.

For simplicity, our simulation supposes that there are three two-valued features, which define eight distinct languages.

(17) \[ +F_1, +F_2, +F_3 \]
\[ +F_1, +F_2, -F_3 \]
\[ +F_1, -F_2, -F_3 \]

etc.

‘Gaps’ occur when certain feature combinations are not attested. Our simulation shows that gaps may arise over the course of time, as the values of two of the features become strongly correlated.

---


3 In fact this must be true in a trivial sense; see Culicover 1999 for discussion.
To take a simple example, if the geographical distribution of \([-F2]\) becomes sufficiently restricted, it may fail to overlap with \([+F1]\). As a consequence, \([+F1]\) and \([+F2]\) become highly correlated. In such a case, some of the languages, namely those with \([+F1, -F2]\), will cease to exist.

Such a situation may occur simply as a consequence of the social structure, and in itself tells us nothing interesting or profound about the relationship between \([+F1]\) and \([-F2]\). I will come back to this point shortly.

For the simulation, we may assume for simplicity that at the outset of the simulation all possible combinations of features are possible (the ‘Tower of Babel’ state). This assumption removes from the simulation any bias in favor of certain feature combinations. If certain combinations fail to exist after some period of time, this fact must be due to social factors, since there are no initial gaps.

Figure 1 shows the random distribution of feature values for three features in a population of 2500 (=50x50). The upper left hand image shows the distinct languages as differences on the gray scale. The other images show the distribution of the + and – values for the three features FIRSTs, SECONDs and THIRDs.

---

4 It is logically possible that the initial population size for various feature combinations determines the final outcome, but we have seen no evidence that this is what is happening our simulations.
The population of each of the eight languages is shown in the histogram in Figure 2. As can be seen, the languages are distributed more or less evenly over the entire population, as would be expected from a randomized assignment of feature values.
I have omitted intermediate steps in the simulation for reasons of time and space. After 69 steps the distribution of languages and features is as in Figure 3.
The histogram in Figure 4 shows the population levels of the eight languages at this point.

![Histogram of languages](image)

**Figure 4. Population of feature combinations after 69 steps**

The loss of languages illustrated in this particular instance of the simulation is not unique. It is a consequence of the particular assumptions made in the simulation about how individuals interact in the network. Running the same simulation under the same parameters yields a different pattern of features and languages each time, but the results are the same.

What happens in the course of clustering is that there is a loss of ‘degrees of freedom’, in the sense that certain possible combinations cease to be realized.

### 3.2.2. Gaps and bias.

Let us now introduce bias into our simulation. The bias may be due to social factors, or to the relatively high intrinsic complexity of a given combination of features compared to the competition. How such complexity is calculated, and
the effect of complexity on learning, must be independently established. The most convincing work along these lines that I know of is that of Jack Hawkins – see for example Hawkins 1994 and Hawkins 2004.

Suppose that a particular combination of features, say \([+F_1, -F_2]\), is less preferred for some reason than the other three combinations of these two features. On every run of the simulation model the results will be more or less the same, in that there will be gaps or immanent gaps in the \([+F_1, -F_2]\) population. It is known that simulations that assume bias in general show a clustering towards the same stable state\(^5\); the strength of the bias determines the predictability of the outcome.

Whether the bias actually causes the particular linguistic features to disappear depends on a number of non-linguistic parameters of the simulation. If the population is small, the bias is more effective than if the population is large, other things being equal. What happens is that speakers with the biased value of the feature switch to the unbiased value of the feature under the influence of the unbiased value in speakers that they interact with. The result of running the simulation with a 10\% bias on one of the feature values (call it ‘+’) on a population of 400 agents is shown in Figure 5.

\(^5\)This is demonstrated in the simulation Sitsim, by Latané, Nowak and Szamrej (Nowak et al. 1990). Kirby 1994 notes the role of bias in change, while Briscoe 2000 has constructed computational simulations of the evolution of language in which biases play a major role.
Figure 5. Simulation after 160 steps, 10% bias on ‘+’ value of feature A, two-valued features (8 languages), 400 agents.

There are two notable characteristics of this simulation. First, only the languages with the ‘+’ value of feature A are still populated. The bias for ‘+’ has guaranteed its rapid success. In addition, it can be seen in this figure that one language is gaining ground over the others. Whether it wins is an open question.

The result of running the simulation with the same parameters on a population of 10000 is shown in Figure 6. Again, the bias for ‘+’ is evident. Because of the larger population, however, the inventory of languages is more stable.
Figure 6. Simulation after 160 steps, 10% bias on ‘+’ value of feature A, two-valued features (8 languages), 10000 agents.

I also ran the simulation on a population of 62500 agents. The population snapshot after 1441 steps is shown in Figure 7, and the distribution of the values for feature A is shown in Figure 8.

---

6 I increased the population to the maximum 62500 that the simulation framework allows in order to minimize the possibility of artefacts due to population size.
Figure 7. Snapshot of simulation after 1441 steps, population = 62500
Figure 8. Snapshot of feature A after 1441 steps

Finally, note that if we increase the bias, say to 100% the languages with the ‘–‘ value of A die out much more quickly, in fact, by 20 steps. See Figure 9.
It turns out that if the agents interact with one another over a broader region of the network, the bias is more effective than if they interact locally. Here we show what happens when we change the interaction distance from 2 to 8. In this case, not only the biased feature but all features cluster much more quickly, and we end up with two languages, not 4.
3.3. Complexity of the mapping

We can use this type of modeling to interpret the relative or total absence of certain logically possible correspondences in natural language, if we can come up with a coherent story about why those correspondences are dispreferred compared to other ways of accomplishing the same sound/meaning mapping. I assume that complexity, however it is ultimately characterized, is at least in part responsible.

I’ve already discussed, one contributor to complexity, which is the transparency, or lack of transparency, of the mapping between syntactic structure and conceptual structure. In case of displacement or dislocation, or what is often called ‘movement’, contiguous portions of the string do not correspond to contiguous portions of the conceptual structure. The structures that result, however they are derived in technical terms, present mismatches for the language learner and the language processor to figure out.

Hawkins 1994 has argued for the view that “words and constituents occur in the orders they do so that syntactic groupings and their immediate constituents
can be recognized (and produced) as rapidly and efficiently as possible in language performance.” Hawkins shows that different constituent orders require different sized spans of a string and corresponding phrase structure in order to determine what the immediate constituents are. The differences “appear to correspond to differences in processing load, therefore, involving the size of working memory and the number of computations performed simultaneously on elements within this working memory.”

We can construct examples in which the mismatch between the order of words and the corresponding hierarchical organization of the phrases is extreme. Let us consider, for instance, the complexity of extracting from a displaced constituent. (18) illustrates.

(18)

---

7One minor concern with the explanatory force of this argument is that we might have expected that human memory would have evolved so as to overcome the problems offered by non-uniform branching. Of course there are many reasons why this would not have happened and it is probably impossible to settle the issue. Shifting the burden of explanation to language acquisition rather than language processing sidesteps this problem, since we probably do not want to attribute adult memory capacity to early learners. See §3.2.2.
Intuitions about complexity suggest that extraction from an extracted constituent is more problematic than extraction from an unmoved constituent. The evidence showing this is due to Postal 1972, who used it as an argument against successive cyclic movement in the Conditions framework of Chomsky 1973.

(19) a. Leslie believes that [a picture of Terry], you would never find ti in a shop like that.
b. *Terry is the person who Leslie believes that [a picture of tj], you would never find tj in a shop like that.

Similarly, examples of the following sort are cited by Wexler and Culicover 1980 as evidence for the Freezing Principle, which blocks extraction from a moved constituent.

(20) a. Whoi did you tell Mary [a story about ti]?
b. *Whoi did you tell tj to Mary [a story about tj]?

The less acceptable examples appear to involve more ‘deformation’ of the canonical structure than do the more accept ones. (This deformation can be characterized in terms of the extent to which syntactic structure reflects conceptual structure.)

It is natural to characterize the complexity of such cases at least in part in terms of the storage requirements that are called upon as the chain that links the wh-phrase and the gap is formed and re-formed and finally interpreted in the course of processing the string. See, for example, Hawkins 1999, Pickering and Barry 1991, Arnon et al. in press and many others for experimental and typological evidence that this relationship is a fundamental one.

Note that extractions such as those in (19) and (20) have been claimed to be possible but marginal. Since judgments of ‘marginality’ do not come labeled with whether they have to do with grammaticality, processing or something else, we may hypothesize that extreme deformation produces complexity (hence processing difficulty). There are subtle differences in unacceptability that we would then expect to account for in terms of a range of factors that contribute to complexity. For examples, examples such as (20b) are judged by some speakers to be grammatical, and examples such as the following are not completely impossible.

(21) ?Terry is the person [of whom]i Leslie pointed out that [such pictures tj], you would never find tj in a shop like that.

What we will expect to find, then, is that form/meaning correspondences that are complex in this sense are rare, will be rejected by speakers, will not constitute the ‘canonical’ way of doing things in a language, but may never the less be exemplified under certain circumstances.

A more general characterization of this view is that grammaticality is the province of the conditions on form per se, and that many grammars (in this sense) are possible. But due to the varying complexity of correspondences between form and meaning, some grammars are more possible than others.
A different sort of problem arises in the case of a construction. In this case, the complexity does not have to do with processing the correspondence, but with the generality of the correspondence. Again, we may say that other things being equal, a more highly specified construction is more complex than a more general construction, or rule. But in this case, there are two considerations at play. One is the complexity, along the lines we have suggested. The other is the on-line processing. While a construction may render the grammar as a whole more complex, it may make processing less complex to the extent that it is a fixed form with a fixed meaning. In this case, the bias may point in the other direction.

4. Grammatical complexity and its implications

The story, then, will incorporate a theory of constructional complexity and a theory of processing complexity, the latter of which bears on both constructions and rules. To model constructional complexity, I suggest that it would be productive to take the approach initially suggested by Chomsky 1965 and implemented in some detail in Chomsky and Halle 1968 as well as Jackendoff 1975. We seek a notation that is more or less transparent in terms of complexity. That is, the complexity of the formal description corresponds to the complexity of the phenomenon.

I do not have time to work through the detail here. Following Simpler Syntax and the work in Construction Grammar, e.g. Goldberg 1995, Goldberg 2006, we may assume that the generalizations about constructions are about their surface properties. We can begin with the notion of a term in the description of a construction, which parallels the notion Structural Description (SD) of Chomsky 1965.

(22) i. A constructional description (CD) is a sequence of terms X, Y, Z, ...
ii. The description … X Y … describes a construction in which X immediately precedes Y.

From this we can define complexity in terms of

- the number of terms,
- the relatedness of the terms to one another,
- the generality of the terms,
- the ordering of the terms, compared with the ordering of terms of the same categories that are not constructional,
- the transparency of the interpretation.

On this view, a CD is more general if it has greater coverage (over lexical items), less strain in terms of the extent to which the construction deviates from the canonical

---

8 That is, following Simpler Syntax, I am setting aside the possibility that syntactic representations are abstract, and that economy conditions stated in terms of these abstract representations play an explanatory role. It is possible that something related to notions of economy in the Minimalist Program (Chomsky 1995) may turn out to correspond to more concrete notions of complexity in terms of processing or learning – see Culicover and Nowak 2002 for some preliminary ideas along these lines.
structure of the language in terms of category membership or constituent order, greater \textit{transparency} to the extent to which it is semantically compositional, and greater \textit{inheritance} to the extent that its properties are inherited from other constructions.

In these terms, a construction that refers to a lexical class is less complex than an otherwise identical construction that refers to a single lexical item in that class; the first has greater coverage. A construction in which the order of constituents conforms to the general order for the language, e.g. \textit{V one's way PP}, is less complex than one in which the words are in a non-canonical order, e.g. \textit{come what may}. Whether one is more stable than the other has to do with relative complexity, but also with frequency, since something that is complex with respect to the overall grammar may persist if it is sufficiently frequent.

Similarly, sluice stranding shows less strain in a language like English that has preposition stranding than it would in a language like French, that does not. Sluice stranding is also semantically transparent, in that the wh-phrase gets a wide scope interpretation in virtue of it being in initial position, and it is interpreted as the complement of the preposition as a consequence of the general pattern of P-stranding in English. Finally, a set of related constructions will be less complex overall to the extent that all of their properties can be accounted for in terms of the family membership – see for example Sag 1997. On the standard view of inheritance, constructions share formal properties, and specialized constructions ‘inherit’ certain properties from other more general constructions. A construction may in principle inherit properties from more than one more general construction; in this sense the constructional properties provide the basic building blocks for more complex constructions.

5. A small case history

I conclude with an concrete example. Consider subject aux inversion (SAI) in English, which is ostensibly the residue of the earlier general V2 property of Old English; modern German is similar enough for purposes of comparison. In Modern English (ME) only auxiliary verbs invert, and only in certain restricted contexts, while general V2 in German applies to all inflected verbs and is triggered by any XP in initial position.

(23) \textit{ME}

a. What are you eating?
b. *What eat you?
c. The banana, I don’t want to eat.
d. *The banana don’t I want to eat.

(24) \textit{German}

a. Was isst du?
what eat you
‘What are you eating?’
b. *Die Banane ich will nicht essen
the banana I want not to-eat
c. Die Banane will ich nicht essen
the banana want I not to-eat
A scenario that I have developed in some detail (Culicover 2007) argues that the transition from V2 to SAI is best characterized in terms of the loss of generality of the V2 rule to the point where it became reinterpreted as a construction. The rule says

\[(25)\quad \text{XP V}^{ [+\text{inflected}]} \text{ NP …}\]

where what follows the verb is ordered by independent principles that determine the articulation of the syntactic structure with focus and discourse, and principles that manage the prosodic structure. (See for example Hawkins 1994, Hawkins 2001, Hawkins 2004, Wasow 1997 and Wasow 2002). In the history of English, the subject NP typically followed the V2. But over time, pronominal subjects began to precede the verb. Baekken 2000:394 gives the following summary of the percentage of XVS and XSV word orders in Early ModE.

<table>
<thead>
<tr>
<th></th>
<th>I: 1480-1530</th>
<th>II: 1580-1630</th>
<th>III: 1680-1730</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>no</td>
<td>%</td>
<td>no</td>
</tr>
<tr>
<td>XSV</td>
<td>4024</td>
<td>80.0</td>
<td>4338</td>
</tr>
<tr>
<td>XVS</td>
<td>1009</td>
<td>20.0</td>
<td>987</td>
</tr>
<tr>
<td>total</td>
<td>5033</td>
<td>100.0</td>
<td>5325</td>
</tr>
</tbody>
</table>

Figure 11. Frequency of the uninverted and inverted word orders (Baekken 2000: 394)

As can be seen, there is a gradual loss of V2 from period I to period III.

We may interpret this change as a reinterpretation of the XVS/XSV alternation from optional variation to a construction. On this scenario, the learner is seeking a syntactic or semantic criterion for V2. Apparent optionality is unstable, on the view that any superficial difference must correlate with something meaningful, either with respect to literal content or discourse structure. XSV is interpreted as a discourse sensitive topicalization construction, while XVS is interpreted as SAI in the context of particular initial X, namely those that correspond transparently to conceptual structure wide scope operators, such as Q and NEG. SAI is thus a construction in the sense that there is a correspondence between a particular configuration and a particular aspect of meaning.

On the other hand, there is considerable evidence that periphrastic do in English began as a very narrow construction, being used with particular verbs, for reasons that are not entirely clear. Ellegård 1953:166 “Of Machyn's 370 ‘do’-instances, 216 involve the verb ‘preach’; the simple verb ‘preach’ occurs only half a dozen times.” and (167) “In Polychronicom there are 816 ‘do’-instances, 243 with ‘slay’ (and no finite ‘slay’ in the past tense), as well as 70 with ‘succeed’, 69 with ‘write’, 19 with ‘eat’, 8 with ‘fight’, 7 with ‘hold’, 7 with ‘appear’ and 4 with ‘add’.” and "we note that verbs or phrases singled out for preferential treatment are not always the same with different writers. We should add to the list Cely Papers /204/ ‘do well understand’, Fitz James ‘appear’, Decaye of England ‘think’, King James Bible ‘eat’." Trudgill et al. 2002 indicate that there was lexical diffusion involving the use and non-use of periphrastic do in negative declaratives: “verbs belonging to the know-group were much slower in acquiring the auxiliary than other verbs.”
Kroch 1989 has analyzed the data in Ellegaard 1953 to demonstrate that do-support developed in each of the contexts in which it currently appears in ModE, at the same rate, but at different stages. There are six contexts distinguished by Ellegård (1953) that Kroch tracks: negative declaratives, negative questions, affirmative transitive adverbial and yes/no questions, affirmative intransitive adverbial and yes/no questions, affirmative wh-object questions, and ‘contact’ do, where do immediately precedes V. Each of these grows in the percentage of cases that do is used vs. the percentage of cases where it could be used, from the period 1400-1425 to the period 1650-1675, with the exception of ‘contact’ do. This use peaked in the mid-16th century and then declined until it was virtually unattested by the end of the 17th century.

These observations suggest that English do-support began as a lexically restricted construction, then split into a set of constructions that became progressively detached from particular lexical items, until it because fully general. Its development as a rule of English is a consequence of its spread through the population, and through the lexicon, until it generalized free of particular items and contexts. Thus it makes the transition from construction to a rule, where the rule may be characterized as follows: “Use do in any context where an auxiliary verb is required but no auxiliary verb is available.”

More generally, it appears to be fruitful to analyze language change in terms of the extension of constructions to rules, which reduces complexity by eliminating idiosyncrasy in the formulation of the constructions. At the same time, it is possible to understand certain changes in terms of the contraction of rules to constructions, which introduces complexities into the form/meaning mapping that can be eliminated by aligning more narrowly defined constructions with more restricted interpretations.

References

Arnon, Inbal, Neal Snider, Philip Hofmeister, T. Florian Jaeger, and Ivan A. Sag. in press. Cross-linguistic Variation in a Processing Account: The Case of Multiple Wh-questions. BLS 32.