Language and Linguistics Compass 2/1 (2008): 59–78, 10.1111/j.1749-818x.2007.00038.x

Variation and Morphosyntactic Theory: Competition Fractionated

David Embick*
University of Pennsylvania

Abstract

The study of the 'dynamic' aspects of language – variation and change – and the development of grammatical theory are often pursued independently of one another. Concentrating on morphosyntax, this article explores connections between these domains that are centered on the notion of 'competition'. When the relationship between *competition for grammaticality* and *competition for use* is articulated, it is clear that *grammar competition* models of variation and change are connected directly to specific theoretical approaches to synchronic grammar.

1 Introduction

There are obvious senses in which theories of morphosyntax do and do not connect with the study of language variation and change. Morphosyntactic theories that investigate grammatical competence do so under familiar idealizations, the result of which is that many of the facets of linguistic behavior that are of interest in the study of variation and change are factored out. Theories of variation and change, on the other hand, assume the categories of morphosyntactic analysis at some level of granularity, but concentrate primarily on the deployment of objects generated by the grammar in situations of use or diachronic development. In a basic way, these are different research programs that differ in terms of the nature of what is to be explained. In the case of morphosyntactic theory, the object of study is the structure of internalized grammars, and ultimately, the sets of constraints on these. In the case of the 'dynamic' aspects of language, the focus is on the different factors that play a role in defining variation and change, such that what is being studied is closer to language use; at least, to a first approximation: in many areas of historical and sociolinguistic inquiry, questions about internal representations are central. As typically approached, these research programs have distinct objectives and methodologies, and it is because of these and other basic differences in the objects of study that any attempt to simultaneously speak of grammatical theory and the dynamic aspects of language is an interdisciplinary endeavor.

At the same time, there are some direct links between these domains. In some cases, the connections exist for reasons that are methodological. In particular, the primary means of acquiring information about the structure of internalized grammars is through forms of data collection in which language use must always be taken into account. There is no direct access to internalized grammars per se; rather, what may be observed is the behavior of individuals who have such grammars. Complicating matters is the fact that – as is standardly assumed – other cognitive systems beyond the grammar play a significant role in determining how speakers behave. Thus, while 'data' may be gathered via a number of techniques, it is not always clear what 'the data' are data about. Rather, it is only in the context of an articulated theoretical model that makes specific claims about how 'cognitive labor' is divided among the different mental systems that are engaged in language and language use that progress toward explanation can be made. In the present context, these points serve as a reminder why so much work in the early development of generative grammar was devoted to methodological clarifications, such as denning key terms like 'grammaticalness', 'acceptability', etc., highlighting the difference between grammatical and memory constraints, and so on. These basic methodological concerns continue to play a central role in current research, a fact that is especially apparent in (i) the need to move beyond the simplest experimental techniques (intuitions) as a primary data source for certain phenomena, and (ii) the (re-)emergence of research programs that take gradience in the data as evidence for gradience in the grammar.[1]

Much could be said, methodologically and otherwise, about these last two points, although this will not be undertaken here beyond some comments on gradience in Section 4. In my view, there is a further question that connects morphosyntactic theory and the study of dynamics, one that is implicated in many current discussions. This central question concerns the status of *competition* in language. By way of illustration, consider the following points and questions:

- Questions about competition are of central importance in syntax and morphology. The discussion of transderivationality, whether in 'syntax proper' (e.g., certain formulations of 'economy'; Chomsky 1991, 1995) or in 'morphological' blocking (Aronoff 1976 and subsequent work), implicates questions about locality and globality that are fundamental to the study of competition. For immediate purposes, the relevant question is, *are there situations in which the grammar generates numerous structures or forms that could potentially express some meaning, with some mechanism then picking out one winner and marking the rest as ungrammatical? That is, is there* ***competition for grammaticality****?*
- In any account of language use, it is true at some level of abstraction that speakers are able to make choices about how to use objects generated by their grammars. The fact that one variant is employed in some situation

means that another variant is not, leading naturally to the idea that distinct variants are competing with one another (cf. Weinreich et al. 1968 for an early variationist formulation). However, unlike what was outlined for *competition for grammaticality* above, the competitors in this kind of competition are all grammatical. *Thus, a kind of* ***competition for use*** *has to be distinguished from competition in the grammar.*

- The existence of 'functionally equivalent' variants (i.e., variation in the sociolinguistic sense) raises a number of questions. *Does the fact that there are sometimes 'doublets' mean that any theory with a Blocking Principle is on the wrong track? Does syntactic variation have implications for the study of synchronic grammar? Or, does the synchronic analysis provide the framework within which the variation can be interpreted? How do* ***competing grammars*** *models of (syntactic) variation and change, argued by Kroch (1994) to be the result of a kind of blocking effect, relate to morphosyntactic theory*?

The goal of this article is to clarify these and related questions about competition, with special reference to the relationship between theoretical and dynamic approaches to morphosyntax. By selecting this particular focus, I am not, of course, making a claim about exclusivity; there are many things to say about linguistic theory and the dynamics of language that are not centered on competition. Rather, I will take as a starting point a specific view of morphosyntax, one that advances explicit hypotheses about the scope of competition in the grammar, and examine some key questions in morphosyntactic dynamics from this perspective.

2 Competition

As a starting point, the questions outlined above point to two types of competition that must be distinguished from one another: competition for grammaticality and competition for use.

At a morphological or morphosyntactic level, a well-known example of a theory involving competition for grammaticality is found in the extremely important analysis of blocking effects formulated in work by Aronoff (1976). Based on the fact that there are *-ity* nouns related to *-ous* adjectives, such as *curious* and *curiosity*, Aronoff develops an account for why some *-ity* nouns like ⋆*gloriosity* are ungrammatical that has competition for grammaticality as a central component. Aronoff observes that the ungrammaticality of ⋆*gloriosity* correlates with the existence of a 'simple noun' *glory*; in the case of *curiosity* and other *-ity* nouns, there is no such simple noun. Aronoff's proposal is that *glory* beats ⋆*gloriosity* for the expression of the meaning 'abstract noun of GLORY'. Thus, according to this theory, although ⋆*gloriosity* is otherwise well-formed, it is ungrammatical, because the 'cell' that ⋆*gloriosity* would occupy is already occupied by *glory*. With, for example, *curiosity*, there is no simple noun, and thus

 Language and Linguistics Compass 2/1 (2008): 59–78, 10.1111/j.1749-818x.2007.00038.x

the *-ity*-affixed form fills the relevant cell. In this theory, it can be seen that the grammar allows multiple potential expressions for a given meaning, with the ultimate winner being determined in a competition that marks the losers as ungrammatical. Put slightly differently, *-ity* forms of all *-ous* adjectives are generated by the rules of the grammar, but whether or not the *-ity* noun is actually grammatical has to be determined by the outcome of a competition.

While Aronoff's case study highlights word/word competition, which fits with ideas about *wordhood* prevalent in certain lexicalist theories, it is possible in principle to extend the scope of competition in the grammar to larger objects, to yield word/phrase (Poser 1992; Hankamer and Mikkelsen 2005; and others) or larger (such as phrase/phrase or clause/clause) 'global' competition (see Bresnan 2001 for a clear formulation; cf. also Andrews 1990 and Kiparsky 2005, 2006, and, for perspectives not centered on 'blocking effects' per se, the papers collected in Barbosa et al. 1998). Generalizing, the term 'competition-based' theory can be used for frameworks that allow competition among different forms or structures, $\mathscr{F}_1 \ldots \mathscr{F}_n$ for the expression of some meaning $\mathscr{M}$. In theories of this type, there are multiple forms competing for the expression of a given meaning; thus, some principle(s) of grammar must determine the winner $\mathscr{F}_w$ of the competition, such that the other competitors are marked ungrammatical. According to standard accounts of this type, there can be only one winner; in the 'paradigmatic' metaphor employed above, this means that there can be only one form for any given cell, that is, for any given meaning.[2]

The set of issues surrounding competition plays an important role in discussions of grammatical architecture. Recent work within the broad heading of the *Minimalist Program* following Chomsky (1993) has highlighted many questions of this type, offering, in effect, programmatic considerations against the kind of globality required by competition-based theories. There are developments in morphosyntactic theory that argue for a similar conclusion on the basis of what competition-based theories predict, in particular, for the conclusion that there is no competition for grammaticality above the level of the morpheme (Embick and Marantz forthcoming). The relevant properties of this approach to competition are summarized in Section 2.

Competition for use is a kind of competition that is not about grammaticality, but, instead, about how expressions are employed by speakers. In any given situation of grammar use, speakers select different options generated by their grammars. Trivially, the choice of one option excludes the use of another in that context (more precisely, in a specific instance of language use). A consequence of this fact is that expressions may compete at the level of use, although each is equally grammatical. To take a rather simplistic example, a speaker might say in a particular instance of use *U* that they were bitten by a *canine*; by doing so, they are unable to use *dog* in *U*, with the result being that various pragmatic effects might be achieved; this is, at the lexical level, the effect of the 'differential'

theory of meaning, of a type most famously associated with Saussure's theory of the sign. Crucially, however, the use of *canine* in such a context does not mean that *dog* is not part of the grammar; it simply means that *dog* was not employed in that particular use of the grammar.

The same considerations exist at a larger structural level as well.[3] This is especially clear in the types of 'information-structure' or 'discourse-related' syntactic alternants. So, for example, the use of *Beans, I like* in some particular situation of use does not render *I like beans* ungrammatical. Each of these options is generated by the grammar, as distinct syntactic objects. The choice in this case is – as in the 'lexical' case – interpretable, in conjunction with other properties of the syntax of topicalization, of course.[4]

Competition for use is of central importance in the study of variation, where it figures importantly in defining exactly when there is variation as opposed to something else; for example, different grammatical options, as in the examples considered above. It is for this reason that early theoretical work that provided foundations for the study of variation like Weinreich et al. (1968) stresses the importance of competition. There is a natural segue here to questions about change. In the examples we have just considered, the competition for use is at the level of expressions that are generated by the same grammar. But the same considerations would extend straightforwardly to expressions that are generated by distinct but closely related grammars; in any given context of use, speakers who command two distinct but closely related grammars will employ one to the exclusion of the other, such that there is a competition for use. This point is clearest in the context of the 'competing grammars' approach to variation and change, which is discussed in Section 4.

3 Morphosyntactic Theory: Some Assumptions

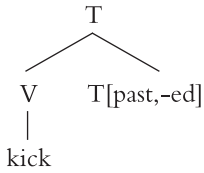
While Aronoff's approach to blocking and others derived from it have led to many important questions, there are reasons to believe that competition for grammaticality is actually quite restricted (more restricted, even, than the word/word competition found in Aronoff's theory). This point has been examined closely in the context of the theory of 'Distributed Morphology' (Halle and Marantz 1993, 1994 and subsequent work), a syntactic, piece-based theory of morphology. In this framework, the morphemes are the nodes that are the primitives of syntactic derivations, and the derivation of complex forms occurs in syntactic structures. The nodes that are manipulated in syntactic derivations are of two types, the *Roots* and the functional morphemes. The *Roots* make up the open-class or 'lexical' vocabulary. They include items such as $\sqrt{\text{CAT}}$, $\sqrt{\text{OX}}$, or $\sqrt{\text{SIT}}$. The functional morphemes are functional categories in the sense familiar from syntactic theory. They consist of non-phonetic features, such as [past] or [pl], or the feature (or features) that makes up the determiner node D of the English definite article *the*. In the PF component of the

grammar, these morphemes receive phonological representations in the process of vocabulary insertion (VI). This process involves vocabulary items like those in (1), which spell out the past tense node T[past] in English; these items compete according to specificity, so that the most highly specified wins. The phonological exponent of the winning VI is inserted into the functional morpheme, as shown in (2):

(1) Vocabulary Items for Past Tense (T[past])

T[past] ↔ -t/__ {$\sqrt{\text{LEAVE}}$, $\sqrt{\text{BEND}}$, …}
T[past] ↔ -∅/__ {$\sqrt{\text{HIT}}$, $\sqrt{\text{QUIT}}$, …}
T[past] ↔ -ed

(2) *kicked*



It is assumed that one VI may apply to a functional node. Thus, the VIs in (1) are in competition, in the sense that the VI inserting *-t* inserts this exponent in the context of, for example, $\sqrt{\text{BEND}}$, and thus prevents the insertion of *-ed*.

An important aspect of the theory of Distributed Morphology is that competition is restricted to VI, which targets individual nodes in the manner described above. This facet of the theory is articulated with reference to blocking phenomena (and compared with alternatives) in Embick and Marantz (forthcoming) (cf. also Embick 2007), where it is argued that there is no blocking between words, nor is there blocking between words and phrases, or larger objects. Thus, *glory* does not block ⋆*gloriosity*; rather, the rules of the language are such that the latter is never generated. This result generalizes to larger objects, so that cases of putative word/phrase blocking ('Poser blocking' named after Poser 1992) are also treated without recourse to word/phrase competition. The only competition for grammaticality that remains in the theory is that seen above in VI, that is, at the level of the morpheme.

The operative principles behind this treatment of blocking effects are central principles of early work in generative syntax. These are as follows:

(3) a. RULES APPLY (RA): Perform a computation when the structural description of the rule is met.
b. No GRADIENCE (NG): Rules are categorical.

A consequence of these assumptions is that a given meaning is not 'potentially expressible' by multiple forms. To take an example that figures

prominently in the literature (Poser 1992; Embick 2007), it is not the case that both *smarter* and *more smart* are generated for the meaning 'comparative of SMART', with the former winning; the former is generated, and the latter is not. The meaning 'comparative of SMART' is defined by a specific set of terminals and features that are the input to the syntactic derivation, and this leads to one morphophonological output.

Taken together, RA and NG make for a specific kind of generative theory, one that has the following property:[5]

(4) Single Output (SO): An input $\mathcal{N}$ to a derivation yields a single output.

From SO, it follows that some input to a syntactic derivation $\mathcal{N}$ has one structural description as its output.[6] By structural description here, I mean <PF,LF> pair. This formulation puts to the side some difficult questions about the nature of the inputs, some of which are addressed below. But it is clear, even at this level of detail, why a theory with the SO property connects directly with certain dynamic matters. If there is one input $\mathcal{N}$ to a syntactic derivation, and we find two distinct forms derived from this $\mathcal{N}$, then there must be distinct grammars at play.

4 Competition and Variation

Theories with SO connect directly with work on morphosyntactic variation and change by Kroch (1989) and others (e.g. Pintzuk 1991), work that proposes a competing grammars (CG) model of variation and change; see also Yang 2002 for a learning model that follows in spirit formulations from Clark 1992 and Clark and Roberts 1993 on genetic algorithms. The central tenet of the CG approach is that certain types of variation are not contained within a single grammar. Rather, there are multiple grammars at play.

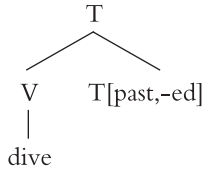
The relationship between SO and CG is direct: in a model with SO, CG is a theorem of the theoretical model, a non-trivial consequence (cf. also Kroch 2001: 720, 722; where the point is made slightly differently). While the relationship between SO and CG is significant, it has some limits; in particular, it is not necessarily the case that all cases of variation are cases of grammar competition. These and other points become clear in the context of specific examples that are important for the dynamics of morphosyntax, the existence of doublets (Section 4.1) and the analysis of movement (Section 4.2) .

4.1 DOUBLETS

As typically defined, doublets are cases in which there appear to be two distinct forms for the same *Root* and combination of features, as shown for the past tense in (5), which has the structure in (6):

(5) dive, PAST dove/dived
dream, PAST dreamt/dreamed

(6) Structure



From the perspective of the theory advanced in Section 3, there cannot be two outputs for the same input. If the syntax generates (6), there is one pronunciation of this object: either *dove* or *dived*, but not both. This is a consequence of SO.

What, then, can the theory say about the coexistence of past tense forms? Assuming that these exist within the output of a single speaker (since variation in a corpus or speech community would be irrelevant to the study of the internalized grammars of individuals), there are two primary options to consider. The first is that the doublets are the outputs of distinct grammars. The other option is of course that SO is incorrect. Each of these requires some comment.

Kroch (1994) implements the former, CG, type of approach, arguing that doublets in the English past tense have a sociolinguistic origin, in this particular case the result of language contact and dialect mixture. Related work by Taylor (1994) examines two different types of change – innovation of an irregular (rare) and innovation of a regular (not so rare):[7] for example, *dived/dove* and *sneaked/snuck* for irregularization, *welk/walked* and *awoke/awaked* for regularization. Importantly, the coexistence of doublets is not stable, whether in the more or less frequent type. Kroch's take on this is that doublets are competing for use, the sense that each doublet comes from a different grammar; that is, any given use of 'past tense of DIVE' requires one of these, such that the other is excluded.

Understanding exactly what the different grammars look like in a CG model requires a theoretical context. Assuming the theory of Section 3, the analysis is that in the *dove*-grammar, DIVE would be on the list associated with the vocabulary item -∅ for the past tense node (as well as on a list of 'readjustment rules' responsible for changing the phonology of the stem); in the *dived*-grammar, DIVE would not be on either list:

(7) a. *dove*-grammar

$$\begin{aligned} &\text{T[past]} \leftrightarrow \text{-t/\_\_ } \{\sqrt{\textsc{Leave}}, \sqrt{\textsc{Bend}}, \ldots\} \\ &\text{T[past]} \leftrightarrow \text{-}\varnothing\text{/\_\_ } \{\sqrt{\textsc{Hit}}, \sqrt{\textsc{Sing}}, \sqrt{\textsc{Dive}} \ \ldots\} \\ &\text{T[past]} \leftrightarrow \text{-ed} \end{aligned}$$

b. *dived*-grammar

$$\text{T[past]} \leftrightarrow \text{-t/\_\_ } \{\sqrt{\textsc{Leave}}, \sqrt{\textsc{Bend}}, \ldots\}$$
$$\text{T[past]} \leftrightarrow \text{-}\varnothing\text{/\_\_ } \{\sqrt{\textsc{Hit}}, \sqrt{\textsc{Sing}}, \ldots\}$$
$$\text{T[past]} \leftrightarrow \text{-ed}$$

In order to sharpen some of what is at issue here, it is worth considering the second type of approach mentioned above, one without SO. In particular, we may consider a 'probabilistic' (PL) approach to the doublet phenomenon, one that makes list membership probabilistic. The idea is as follows: what is required in order to produce irregular allomorphy is access to a list, as listedness is what controls insertion of non-default allomorphs (and stem-changing rules as well). The failure to access a list has been implicated in other domains (e.g., allomorphic 'overregularization' phenomena in acquisition of the type examined in Marcus et al. 1992 and related work). What is being considered, then, is a version of the theory presented in Section 2 in which we have instead of (7a) or (7b) something like (8):[8]

(8) Vocabulary Items for Tense: (PL) Version

$$\text{T[past]} \leftrightarrow \text{-t/\_\_ } \{\sqrt{\textsc{Leave}}, \sqrt{\textsc{Bend}}, \ldots\}$$
$$\text{T[past]} \leftrightarrow \text{-}\varnothing\text{/\_\_ } \{\sqrt{\textsc{Hit}}, \sqrt{\textsc{Sing}}, \mathcal{P}_i \sqrt{\textsc{Dive}} \ldots\}$$
$$\text{T[past]} \leftrightarrow \text{-ed}$$

Here $\mathcal{P}_i$ is the probability at which *dove* appears as the past tense of *dive*. Within this grammar, there is essentially a coin-toss. If the situation is one in which $\sqrt{\text{DIVE}}$ is on the $\varnothing$ list, then *dove* results; if not, then we find *dived*.[9] The intuition here is that the grammar per se is not 'gradient' in any strong sense, as there is nothing about the syntax that is affected by probabilistic listing, which is restricted to allomorphic selection.

At the level of what is produced by the grammar(s), it is not clear how to distinguish the CG approach from the PL approach. Each allows for the derivation of both *dived* and *dove*, and, although they make very different claims about how each of these forms come about, it is not obvious that there is anything at the theoretical level that distinguishes the two; at least, there are no convincing arguments that I am aware of.

There is, on the other hand, a clear difference between these accounts when we take modularity into consideration. Distinguishing between competition for grammaticality and competition for use in the manner of Section 2 results in a modular theory in which questions about grammaticality are separated from questions about how the grammar is employed.

This needs to be made explicit:

(9) Modularity Assumption (MA): *Grammar* and *language use* are modularly distinct.

Here, the CG and PL theories under consideration diverge. While the CG approach allows for (9) to be retained, there are cases in which a PL treatment along the lines sketched above in (8) forces a blurring of the boundaries between grammar and use. This point is clearest in cases in which the distribution of variants is conditioned by external (i.e., sociolinguistic) factors. While this is not obviously the case with the past tense doublets above, relevant examples are not difficult to find. Take, for an additional example, the case of the present participial Asp[pres] exponent, which (simplifying) alternates between *-ing* and *-in* (there is a substantial literature detailing these effects, following Fischer 1958 and Labov 1966). Because the phonological exponent of a VI has to be memorized, one could imagine extending the PL approach to this cases as well, as in (10):

(10) Asp[pres] $\leftrightarrow$ $\mathscr{P}_i$-in
Asp[pres] $\leftrightarrow$ $\mathscr{P}_j$-ing

As in the example with T[past] above, the idea is essentially that in a derivation with the node Asp[pres], there is a coin-toss in the phonological component, and this determines which allomorph is selected.

The $\mathscr{P}$-values in (10) must be tailored to a specific set of surface distributions. As is well-known, the distributions in this case are conditioned by a number of external factors (putting aside internal ones), in particular socioeconomic class and contextual style. The big question is to how to get this sort of information to determine the probabilities: that is, how to make the $\mathscr{P}$-values that are by hypothesis in the grammar be determined by factors that are 'external' in nature. In (10), the fact that the different allomorphs are correlated with distinct conditions of use is not encoded in the coin-toss procedure. This clearly is insufficient, so the theory has to be augmented. The only obvious means of doing so would be to build the sociolinguistic contexts into the probability calculation. There are different ways of doing this, but they are all non-modular in the relevant sense.

The modularity questions raised here are not new. Since variable phonological rules were originally proposed (cf. among others, Labov 1969 and Cedergren and Sankoff 1974), an important question has been how to make the grammar sensitive to external factors (see, for example, Fasold's (1991) discussion, situated in the context of a premature postmortem for variable rules). Questions of this type are becoming increasingly more important to contemporary discussions, as 'gradient' conceptions of grammar have come back into fashion (e.g. Bod et al. 2003). The important point for the purposes of this article is that the PL analysis of (10) requires

the lines between grammar and use to be blurred in some fashion, and this is, of course, a consequence that must be recognized both by those who make this move, and those who argue vehemently against it. See Labov (2001; 28) for some important observations about internal vs. external factors.

The CG approach does not modify the grammar in order to account for correlations with external factors, nor does it modify the Modularity Assumption highlighted above. In a CG model, there may or may not be a sociolinguistic effect on the distribution of variants. If there is, then which grammar a speaker chooses to employ is affected by particular conditions of use. This is not a particularly profound observation; the CG model is, in a sense, intended in part to maintain a sharp distinction between grammar and use. But it serves to highlight the point that if such a distinction is to be maintained, a competing grammars approach (or something like it) is forced.

In summary, both theories have something to say about the doublet phenomenon, but they differ in terms of the Modularity Assumption (9), and, thus, in terms of how exactly the effects are accounted for. Ultimately, it has to be asked what the benefit of putting probabilities in the grammar is, whether in allomorphy or beyond. We see in the present case considered above that there is no extra empirical benefit to doing so: both the CG and PL approaches can account for surface distributions, they simply localize the effect in different cognitive systems. In general, it seems quite likely that in any case of this type, it would be possible to weight the rules in such a way as to derive the correct surface distributions. But all this shows is that an approach whose primary focus is accounting for surface distributions alone is of little interest, since, after all, that is what we start with.

4.2 MORPHOSYNTAX AND MOVEMENT

A great deal of work in historical (morpho-)syntax has concentrated on changes in movement rules; movement rules relating to head movement and affixation in particular. Consider, for example, the loss of verb movement in English: the process by which English lost movement of V from within the VP to a higher functional head (given here as T for T(ense)). The different stages of the language with and without movement are detectable with reference to the position of adverbs, which are assumed to be VP adjoined (examples cited in Kroch 1989):

(11) a. Quene Ester [$_{T}$ looked [$_{VP}$ never with swich an eye]].
(Chaucer, *Merchant's Tale,* 1744)
b. Queen Ester [$_{VP}$ never looked with such an eye].

In Middle English (11a), the verb appears adjoined to the T head to which it has moved; as such, the verb precedes the VP-adjoined adverbial. In Modern English, this movement does not occur for main verbs.

Rather, Tense is LOWERED to V (cf. Embick and Noyer 2001 for a proposal), such that the adverb appears before the verb linearly.[10]

The central question for this overview is how competition for use and competition for grammaticality are related in the domain of movement. In this domain, Kroch (1994) attempts to reduce the CG theory to the factors responsible for blocking effects. The particular intuition is based on the motivation for blocking effects presented in certain versions of lexical phonology (Kiparsky 1982, 1983): the idea that distinct, functionally equivalent forms cannot exist.[11]

Kroch assumes a theory in which movement is feature driven, in the sense of Chomsky (1993). In such a theory, the locus of parameterization is the system of functional heads. When a functional head like Tense is 'strong'(+), it induces overt movement; when it is weak(–), there is no overt movement. Kroch (1994) proposes that one functional head – for the example above, T– cannot be both STRONG and WEAK in the same language. This restriction on the inventory of syntactic primitives is hypothesized by Kroch to result from blocking: T+ and T– would be 'functionally equivalent', and therefore they should block one another. Because it regulates the inventory of syntactic primitives irrespective of how they are combined, let us refer to this principle as a type of INVENTORY CONSTRAINT (IC):

(12) INVENTORY CONSTRAINT: Functional head X cannot be both strong+ and weak– in the same language, because X+ and X– block each other.

The effect of the particular IC in (12) proposed by Kroch is to make grammar competition a consequence of the way that the syntax works. In any case in which 'the data' contained verb movement and no verb movement for the same input, this could only be the consequence of having T+ and T–. Since these two heads cannot coexist in the same grammar according to Kroch's blocking-motivated IC hypothesis, the pattern in the data must reflect the existence of two distinct and competing grammars, that is, one with T+ and one with T–. According to this view, the CG model of variation is a consequence of a component of the synchronic grammar (i.e., of IC) at least in the case of variation in feature-driven movement. However, this hypothesis derives CG in a different way from what was discussed in Section 3 above, where CG was shown to follow from SO.

4.3 COMPETITION, INVENTORY CONSTRAINTS, AND SINGLE OUTPUT

We have seen above two different principles that force a CG type theory: SO and Kroch's blocking hypothesis, which centers on a particular IC. That these principles do different work is clear from the fact that SO says nothing about what heads happen to exist in a particular language. It says only that an input to a derivation can have a single output. A theory with

Kroch's IC, on the other hand, goes beyond this by specifying what properties inputs to derivations can have, by preventing strong and weak functional heads that are otherwise identical from co-occurring in the first place.

An obvious question is how much empirical support there is for IC. While some apparent counterexamples are probably in fact only apparent,[12] there is reason to believe that there are other cases that pose difficulties.

Working through any particular example of this type requires detailed assumptions that go well beyond the scope of the present discussion. However, in order to illustrate the issues (and introduce some questions about what is at stake), I will outline a simplified argument of the relevant type in the remainder of this subsection.

Consider, to begin with, the pattern of 'subject–aux' inversion in English questions, also known as T-to-C movement. This is a process that moves the Tense node to C when C has the feature [wh].[13] In the matrix context, the C[wh] induces movement of both the *wh*-expression and T, as in (13a). In embedded questions, on the other hand, *wh*-movement occurs, but there is no T-to-C movement (13b):

(13) a. i. What did John play *t*?
ii. ⋆What John played *t*?
b. i. ⋆Mary asked [what did John play *t*]
ii. Mary asked [what John played *t*]

Using '+/–' for the trigger of T-to-C movement, what we see is that English evidently has both C[wh+], and C[wh–] (see Pesetsky and Torrego 2001 for recent discussion).[14] The former occurs in matrix clauses, the latter in embedded clauses. If in fact the C nodes in the embedded and matrix contexts are the same, this is the situation that the IC rules out. It is not clear how this conclusion could be avoided, short of holding that the embedded C and the matrix C are fundamentally different objects.[15] Determining whether or not the heads are in fact the same is difficult. For the purposes of the discussion here, I will assume that they are identical, however, as the point is to further the discussion of the empirical issues related to various inventory constraints.

In order to derive the correct pattern, it must be encoded somewhere in the grammar that C[wh+] is found in matrix clauses, while C[wh–] is employed in embedded clauses. The pattern cannot be stated in a theory with IC, according to which positive and negative values for the same head should never coexist. Apparently what is required is a theory that can enforce something like (IC2):

(14) (IC2): Z+ and Z– may both occur, but not in the same contexts.

Another way of putting this is that the ± value of an individual head is too small to account for the ways in which languages differ; aspects of the broader syntactic context must be included in the formulation of such statements as well. So, for example, the 'contextual' effects on C-values

 Language and Linguistics Compass 2/1 (2008): 59–78, 10.1111/j.1749-818x.2007.00038.x

in English shows variation as well, as is clear from examples in which there is movement in embedded clauses, possible in Belfast English (Henry 1995):

(15) a. She asked [who had I seen].
b. They wondered [what had John done].
c. I wondered [where were they going].

Clearly, one cannot just say that while speakers of Belfast English have to acquire C[wh+], speakers of other varieties of English have to acquire both C[wh+] and C[wh–]. In addition to this, learners in the latter case must learn where the different C heads are employed. Framed slightly differently, the important point is that while the mechanical effects of parameterization can be reduced to properties of heads, this does not exhaustively describe how languages differ in terms of movement. Something further concerning the distribution of heads – presumably something encoded via selection – must be acquired as well. Thus, the locus of parametric differences between languages is, in some sense, 'larger' than the individual head, a point that raises many questions of its own (see, for example, Clark 1994 and references cited there for discussion of the locality domains over which parametric differences are be manifested). A more detailed discussion of these matters could examine connections between the relevant domains and the cyclic domains at the heart of 'phase theory' à la Chomsky (2001) and related work.

In summary, while Kroch's (1994) proposal speaks of competing grammars as being the result of blocking, it appears that blocking as typically conceived is not necessarily at issue. Rather, what the proposal highlights is the fact that, independent of the question of synchronic variation, a theory of inventory constraints is necessary, and, in addition to this, that certain ICs can have implications for how synchronic variation is understood.[16] All of these matters are connected closely with core theoretical questions, in particular those concerning the nature of what is input to syntactic derivations and the nature of optionality; while the details cannot be investigated further here, it should be noted that this close connection between theoretical and dynamic issues is exactly what one would hope for in a truly interdisciplinary theory.

4.4 FURTHER DIMENSIONS

There are two primary questions that will shape future investigations into the dynamics of competition, beyond the syntactic concerns with ICs discussed above. The first concerns the scope of the CG model. While the discussion immediately above focusses on syntactic movement, there are of course many instances of variation in other domains as well. A pressing question, which goes beyond the confines of this discussion, is whether, for example, phonological and phonetic variation should be treated in CG terms.

The second question, broached above in the discussion of doublets, is how CG models fare in comparison with 'gradient' or 'probabilistic' theories of competence; theories that, in the terms employed above, abandon NG. While it is not to be doubted that statistical or gradient concerns play a role somewhere in language in the broad sense, a further question is of course where in the cognitive system such effects are to be located: this is a modularity question.

One has to be wary here of certain confusions, concerning what the different theories under discussion are trying to explain. Theories with the Modularity Assumption (9) separate grammar from language use, and, in the 'classic' generative sense, aim to explain properties of the former. At the same time, they do not deny that there are patterns of language use (probabilistic or otherwise) that are worthy of explanation; rather, they say that grammar and use are dealt with via distinct cognitive mechanisms that should not be conflated. The claim that 'generative theories' with MA and NG fail to account for such effects might therefore be true, but it is not probative; it might, in some sense, be like saying that the cognitive neuroscience of vision does not account for the aesthetics of sculpture, or whatever else the visual system might be employed for. What is required is a demonstration that the probabilistic effects are crucially part of the grammar per se, not part of some other system. Merely pointing to statistical patterns and claiming that they are not accounted for by some treatment or other falls far short of this.[17]

In the end, theories that hold to the MA make a sharp distinction between (i) what is to be explained (e.g. the grammar, language use, etc.), and (ii) the evidence that is used to arrive at (i) (judgments, statistical patterns, etc.). Some accounts of the probabilistic type are quite explicit about the conflation of these domains, and assume explicitly that because 'the data' are 'gradient', the grammar must be gradient as well. For example, Hay and Baayen (2005) conclude that the 'categorical distinction between simple words . . . and complex words' has to be abandoned because '. . . people's behavior in experimental tasks is anything but categorical.' Although much could be said about the particular points raised in that paper (and related work), all such discussions have to begin with the fact that the reasoning from gradience in some behavioral measure to gradience in the grammar simply does not follow. It is a possibility, but, of course it is also possible that the gradience arises elsewhere. In any experimental context probing well-formedness, the theory of the task identifies a number of cognitive systems that have some influence on subjects' responses, and 'the grammar' in the narrow sense is only one of these. Thus jumping from 'gradient data' to 'gradient grammar' in the absence of any strong argument for doing so does not exactly clarify the real issues.

In the final analysis, these questions have to be settled empirically. In terms of grammatical models of competition, there seems to me to be no

reason to move away from theories with 'classic' generative properties like that in Section 2, which has SO and, thus, CG as a consequence. Correspondingly, it is important to note that in the diachronic domain, empirical arguments in favor of CG have been provided as well. See, for example, the discussion of the *Constant Rate Effect* and its relation to the CG model in Kroch (2001); if these considerations on the right track, then there are clear empirical arguments in favor of the CG model, to the extent that it provides a better explanation of such effects than surface-based alternatives. It would, however, take another review to do justice to the *Constant Rate Effect* and compare it to alternative explanations.

5 Concluding Remarks

The notion of competition is central to both theoretical and variationist research programs. In addressing morphosyntactic variation, modular theories distinguish sharply between competition for grammatically and competition for use. There are reasons for believing that the former type of competition is highly restricted, such that many 'blocking effects' analyzed in some theories via competition are in fact the result of the workings of a generative theory with the SO property. As a general point, theories of this type have CG as a consequence, an important consequence that is being investigated empirically, especially in contrast with 'probabilistic' alternatives. While the field seems to be moving toward a comparison of categorical and probabilistic models of language, it is important to stress, first, that many of the relevant issues center on the notion of competition; and second, that, because in many cases plausible accounts of some phenomenon can be given in either categorical or probabilistic terms, empirical arguments are required for determining which view is correct. From the fact that there exist variable patterns in language, it can only be concluded that some part of the many integrated cognitive systems that make up human language use shows gradience somewhere. Any further conclusions require detailed assumptions about mechanisms and architecture of the type outlined above.

Short Biography

For comments I would like to thank Rajesh Bhatt, Tony Kroch, Marjorie Pak, Don Ringe, and the participants in my Fall 2006 seminar at the University of Pennsylvania.

Notes

* Correspondence address: David Embick, Department of Linguistics, 619 Williams Hall, University of Pennsylvania, Philadelphia, PA 19104-6305, USA. Email: embick@ling.upenn.edu.

[1] For (1) see Labov (1996), Schütze (1996), and Marantz (2005) for views that range from sociolinguistic to theoretical linguistics, psycholinguistics, and the cognitive neuroscience of language; for (2) the papers in Bod et al. (2003) and the discussion later in the paper.
[2] Thus, one possible approach to variation is to relax this assumption in some fashion, although it is not clear what convincing evidence for this move would look like; see the discussion below.
[3] As smaller levels too, for cases with morphological or phonological alternants are employed; see Section 3.
[4] In this case, as in many other aspects of interpretation, the crucial question is: what aspects of the 'meaning' of an expression in the broad sense are encoded in the grammatical representation in the strict sense, and which arise via semantic or pragmatic mechanisms that take into account (among other things) other potential expressions? This question is extremely complicated, and one result of this fact is that there are cases in which it is not clear at present what kind of competition is at issue.
[5] The claim is not that SO is restricted to theories without competition; there are many different types of theory that can have SO as a component. In a competition-based theory with SO, all competitors save one must be eliminated.
[6] Another way of putting this is that given a particular starting point for a derivation, there is no optionality in what results; at the same time, there may be optionality in the broad sense depending on what derivations are allowed to begin with; see Section 4. For related discussion of optionality, see also Roberts (2007: 305sqq).
[7] These are assumed to be cases in which there is no differentiation of meaning or morpho-syntactic structures associated with the different allomorphs; see Kroch (1994) for past tense examples of this, Embick (2003) for participial instances.
[8] In terms of the theory of Section 2, making list-access probabilistic works better than, for example, doing the same with the blocking mechanism; see Yang (2002) for discussion of 'stochastic blocking' along these lines. When taken within a piece-based framework 'stochastic blocking' would predict, evidently, a failure of one type of piece to block another, as blocking results from the assumption that only one VI may occur in any given node. Thus, making this effect stochastic would predict the existence of double marking, as in, for example, **ben-t-ed*, at whatever $\mathscr{P}$ blocking is supposed to fail for. For more familiar doublet cases, this would be **dov-*$\varnothing$*-ed*, **learn-t-ed*, etc.
[9] We would have to correlate this with the readjustment rule as well. This could be done by having that rule sensitive to the $\varnothing$-exponent, in the case at hand. However, we do this, the point is that this grammar generates both *dived* and *dove* out of this one grammar.
[10] Strictly speaking, the word order in (11b) is also compatible with the adverb being adjoined higher as it is in *John never can finish all of his assignments*, where the modal *can* is in T.
[11] This is a motivation for blocking that derives from the works cited. The idea that blocking relationships can be defined in terms of functional equivalence is potentially problematic, depending on its scope. For example, Poser (1992) effectively criticizes 'pragmatic' theories of blocking, raising objections not addressed in recent revivals of 'Gricean' blocking, such as Williams (2007).
[12] For example, consider the fact that in English auxiliaries move to T, while main verbs do not; this would apparently require both T+ and T–. In this case, however, one could argue that the auxiliaries are a subcategory of *v*, in a way that makes the T+ relevant for Aux-to-T movement distinct from the 'normal' T found in clauses without auxiliaries. Presumably, this could be encoded via selection, although implementing this type of analysis might not be trivial.
[13] I am abstracting away from other cases of T-to-C movement (with negative expressions, counterfactuals, etc.) in this discussion.
[14] As pointed out by a reviewer, there are many other assumptions at play here, such as the assumption that head-movement targets C in both of these environments.
[15] More radically, one could argue that it is head movement that is at issue here. Within certain theories, for example, that of Chomsky (2001), this would not be a syntactic problem in the strict sense, as that theory puts head movement in the PF component of the grammar. Within that theory, however, the property of forcing overt movement (i.e. the 'EPP-property' or its equivalent) may or may not be associated with the same head. For example, the 'strong phase' defining *v* heads must induce movement of *wh*-expressions prior to the movement of such

expressions to C. This means that there are semantically identical *v* heads that differ only in EPP properties. Thus, the same point could be made with reference to phrasal movement.

[16] As noted above, SO forces a CG analysis, but says nothing about what the inputs to derivations are. If the inputs to derivations can have the same heads in more than one variety, then a theory can have the SO property while still allowing what appears to be 'two ways of saying the same thing' (i.e. optionality). Something like this appears to be at the center of Adger and Smith's (2005) proposal that certain functional heads may differ only in terms of an 'extra' uninterpretable feature.

[17] The fact that there must be some additional empirical or theoretical connections in '(CG) versus probabilistic grammar' discussions is not always appreciated. For example, Henry (2002: 273) asserts that '. . . there is a considerable difference between a person having two grammars, and a single grammar which admits optionality'; this is clearly true, but Henry offers no argument, empirical or theoretical, to distinguish these different approaches, and concludes nevertheless that '. . . a better characterization [of variation] seems to be that individual structures/parameter settings are variable, rather than that there are actually separate grammars.' (2002: 274). This is not to say that Henry's work falls into the same category as the papers in, for example, Bod et al. (2003), however. In fact, Henry's analyses are closer in spirit to the assumptions motivating the CG approach. My point is that much caution is required in reasoning about the relationship between variation and grammar.

Works Cited

Adger, David, and Jennifer Smith. 2005. Variation and the minimalist program. Syntax and variation: reconciling the biological and the social, ed. by Leonie Cornips and Karen P. Corrigan, 149–78, Amsterdam, The Netherlands/Philadelphia, PA: John Benjamins.

Andrews, Avery. 1990. Unification and morphological blocking. Natural Language and Linguistic Theory 8.507–57.

Aronoff, Mark. 1976. Word formation in generative grammar. Cambridge, MA: MIT Press.

Barbosa, Pilar, Danny Fox, Paul Hagstrom, Martha McGinnis, and David Pesetsky. 1998. Is the best good enough? Optimality and competition in syntax. Cambridge, MA: MIT Press.

Bod, Rens, Jennifer Hay, and Stefanie Jannedy. 2003. Probabilistic linguistics. Cambridge, MA: MIT Press.

Bresnan, Joan. 2001. Explaining morphosyntactic competition. Handbook of contemporary syntactic theory, ed. by Mark Baltin and Chris Collins, 1–44. Oxford, UK: Blackwell.

Cedergren, Henrietta, and David Sankoff. 1974. Variable rules: performance as a statistical reflection of competence. Language 50.333–55.

Chomsky, Noam. 1991. Some notes on economy of derivation and representation. Principles and parameters in comparative grammar, ed. by Robert Friedin, 417–54. Cambridge, MA: MIT Press.

——. 1993. A Minimalist program for linguistic theory. The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger, ed. by Ken Hale and Samuel Jay Keyser. Cambridge, MA: MIT Press.

——. 1995. The minimalist program. Cambridge, MA: MIT Press.

——. 2001. Derivation by phase. Ken Hale. A life in language, ed. by Michael Kenstowicz, 1–52. Cambridge, MA: MIT Press.

Clark, Robin. 1992. The selection of syntactic knowledge. Language Acquisition 2.83–149.

——. 1994. Kolmogorov complexity and the information content of parameters, Technical Report 94–17. Philadelphia, PA: Institute for Research in Cognitive Science, University of Pennsylvania.

Clark, Robin, and Ian Roberts. 1993. A computational model of language learnability and language change. Linguistic Inquiry 24.299–345.

Embick, David. 2003. Locality, listedness, and morphological information. Studia Linguistica 57.143–69.

——. 2007. *Blocking effects* and analytic/synthetic alternations. Natural Language and Linguistic Theory 25.1–37.

Embick, David, and Alec Marantz. forthcoming 2008. Architecture and blocking. Linguistic Inquiry 39.
Embick, David, and Rolf Noyer. 2001. Movement operations after syntax. Linguistic Inquiry 32.555–95.
Fasold, Ralph W. 1991. The quiet demise of variable rules. American Speech 66.3–21.
Fischer, J. L. 1958. Social influences on the choices of a linguistic variant. Word 14.47–56.
Halle, Morris, and Alec Marantz. 1993. Distributed morphology and the pieces of inflection. The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger, ed. by Kenneth Hale and Samuel Jay Keyser, 111–76. Cambridge, MA: MIT Press.
——. 1994. Some key features of distributed morphology. Papers on phonology and morphology, ed. by Andrew Carnie, Heidi Harley and Tony Bures, vol. 21, 275–88. Cambridge, MA: MIT Working Papers in Linguistics.
Hankamer, Jorge, and Line Mikkelsen. 2005. When movement must be blocked: a response to Embick and Noyer. Linguistic Inquiry 36.85–125.
Hay, Jennifer, and Harald Baayen. 2005. Shifting paradigms: gradient structure in morphology. Trends in Cognitive Sciences 9.342–248.
Henry, Allison. 1995. Belfast English and Standard English: Dialect variation and parameter setting. Oxford, UK: Oxford University Press.
——. 2002. Variation and syntactic theory. The handbook of language variation and change, ed. by J. K. Chambers, Peter Trudgill and Natalie Schilling-Estes, 267–82. Oxford, UK: Blackwell.
Kiparsky, Paul. 1982. Lexical morphology and phonology. Linguistics in the morning calm: Selected essays from SICOL-1981, ed. by Linguistic Society of Korea. Seoul, South Korea: Hanshin.
——. 1983. Word formation and the lexicon. Proceedings of the 1982 Mid-America Linguistics Conference, ed. by F. Ingemann, 3–22. Lawrence, MA: University of Kansas.
——. 2005. Blocking and periphrasis in inflectional paradigms. Yearbook of Morphology 2004, ed. by Geert Booij and Jaap van Marle, 113–35. Dordrecht, The Netherlands: Springer.
——. 2006. Grammaticalization as Optimization. MS. Stanford University.
Kroch, Anthony. 1989. Reflexes of grammar in patterns of language change. Language Variation and Change 1.199–244.
——. 1994. Morphosyntactic variation. Papers from the 30th Regional Meeting of the Chicago Linguistics Society: Parasession on Variation and Linguistic Theory, ed. by Katharine Beals, 180–201. Chicago, IL: Chicago Linguistic Society.
——. 2001. Syntactic change. The handbook of contemporary syntactic theory, ed. by Mark Baltin and Chris Collins, 699–729. Oxford, UK: Blackwell.
Labov, William. 1966. The social stratification of English in New York City. Washington DC: Center for Applied Linguistics.
——. 1969. Contraction, deletion and inherent variability of the English copula. Language 48.773–818.
——. 1996. When intuitions fail. Papers from the parasession on theory and data in linguistics: April 11–13, 1996, ed. by L. McNair, K. Singer, L. Dolbrin and M. Aucoin, 76–106. Chicago, IL: Chicago Linguistic Society.
——. 2001. Principles of linguistic change, vol. 2: social factors. Oxford, UK: Blackwell.
Marantz, Alec. 2005. Generative linguistics within the cognitive neuroscience of language. Linguistic Review 22.429–45.
Marcus, Gary, Steven Pinker, Michael Ullman, Michelle Hollander, T. John Rosen, and Fei Xu. 1992. Overregularization in language acquisition (Monographs of the Society for Research in Child Development). Chicago IL: University of Chicago Press.
Pesetsky, David, and Esther Torrego. 2001. T-to-C movement: causes and consequences. Ken Hale: A life in language, ed. by Michael Kenstowicz. Cambridge, MA: MIT Press.
Pintzuk, Susan. 1991. Phrase structure in competition: variation and change in Old English word order. Doctoral dissertation, University of Pennsylvania.
Poser, William J. 1992. Blocking of phrasal constructions by lexical items. Lexical matters, ed. by Ivan Sag and Anna Szabolsci, 111–30, Stanford, CA: CSLI.
Roberts, Ian. 2007. Diachronic syntax, Oxford, UK: Oxford University Press.
Schiitze, Carson T. 1996. The empirical basis of linguistics: grammaticality judgments and linguistic methodology. Chicago, IL: University of Chicago Press.

Taylor, Ann. 1994. Variation in past tense formation in the history of English, University of Pennsylvania Working Papers in Linguistics, vol. 1, ed. by Roumyana Izvorski et al., 143–59.
Weinreich, Uriel, William Labov, and Marvin I. Herzog. 1968. Empirical Foundations for a theory of language change, Directions for historical linguistics, ed. by W. Lehmann and Y. Malkiel, 95–195. Austin, TX: University of Texas Press.
Williams, Edwin. 2007. Dumping lexicalism. Oxford handbook of linguistic interfaces, ed. by Gillian Ramchand and Charles Reiss. Oxford, UK: Oxford University Press.
Yang, Charles. 2002. Knowledge and learning in natural language. Oxford, UK: Oxford University Press.

