

DISCUSSION NOTES

On stochastic grammar*

BRADY CLARK

Northwestern University

1. INTRODUCTION. Newmeyer (2003) has argued against models of mental grammars that incorporate probabilistic information (henceforth, *STOCHASTIC GRAMMAR*). In the discussion that follows, I review N's arguments against stochastic grammar and attempt to show that they do not stand up to scrutiny.

More generally, what follows is a defense of the *INHERENT VARIABILITY* tradition of modeling linguistic variation.¹ The inherent variability tradition includes the variable rules approach (see Labov 1969 and references in Paolillo 2002), classification and regression tree analysis (Ernestus & Baayen 2003), analogical modeling of language (Skousen 1989), generalized linear models (see references in Manning 2003), various versions of optimality theory (e.g. stochastic optimality theory (Boersma 1998, Clark 2004), partial ordering (Anttila 1997), floating constraints (Nagy & Reynolds 1997)), extensions of head-driven phrase structure grammar (Bender 2001), and extensions of the principle and parameters framework (Yang 2003). A guiding assumption of work in this tradition is that mental grammar accommodates and generates variation, and includes a quantitative, noncategorical, and nondeterministic component (Weinreich et al. 1968, Bender 2001).

Before I turn to N's arguments, I must mention some terminological assumptions. Throughout, when I use the term *MENTAL GRAMMAR*, I mean a linguistic system existing in the mind of an individual speaker. When I use the term *GRAMMAR MODEL*, I mean a theory of the mental grammar of an individual.

2. METHODOLOGICAL ISSUES. N's (2003:696) first argument against stochastic grammar is methodological in nature. Noting that the corpora that advocates of stochastic grammar have typically drawn probabilities from encompass data from a wide range of speakers (e.g. the Switchboard Corpus (Dick & Elman 2001), the *New York Times*, etc.), N asks 'how could usage facts from a speech community to which one does not belong have any relevance whatsoever to the nature of one's grammar' (Newmeyer 2003:696).

2.1. ON THE USE OF CORPORA. This is a fair question, but its pointedness is weakened when one considers what linguists use corpora for. In practice, large corpora are drawn upon to develop descriptions of the language use of a wide range of speakers, as in statistical natural language processing (Manning & Schütze 1999:7). The central question in this type of approach is whether the corpora that are available are representative. This question is addressed by using statistics to handle finite samples of potentially infinite sets of data (Brew & Moens 2000).

Is it reasonable to use frequency asymmetries in corpora to justify theories of *INDIVIDUAL* mental grammars? As a matter of convenience and intended applications (e.g.

* I would like to thank Joan Bresnan, Brian Joseph, Paul Kiparsky, K. P. Mohanan, Fritz Newmeyer, Richard Oehrle, Peter Sells, Elizabeth Traugott, and Annie Zaenen for their comments on earlier versions of this discussion note.

¹ See also Abney 1996, Manning & Schütze 1999, Wasow 2002, and Manning 2003.

multi-user speech-to-text systems), work in statistical natural language processing typically does not separate data from different individuals. Consequently, descriptions of language use drawn from multi-individual corpora may seem irrelevant to grammar models. However, if we make the idealization that speakers share the same mental grammar, we can use the frequency of data of multi-individual corpora as a model of individual variation. Without this idealization, we could not begin to understand what different speakers have in common, for example, within a given speech community. Further, it is an accidental fact, not an essential one, that present-day corpora are not tailored to single individuals, in the sense that they contain all the external linguistic evidence a speaker might encounter or be exposed to in the course of learning. Present-day corpora serve as an adequate surrogate for individually tailored longitudinal collections of that sort. These corpora may not be a perfect substitute, but they are the best we can manage currently.²

One possible objection is that by mixing data from different individuals together in a large data set, evidence relevant to the investigation of the mental grammar of particular individuals is potentially obscured (Mohanani 2003, Newmeyer 2003). For example, in a situation where some individual mental grammars have pattern X (e.g. verb-object) and other individual mental grammars have pattern Y (e.g. object-verb), combining data from these two types of linguistic systems results in apparently random free variation (Mohanani 2003).³

It is often methodologically necessary to look at groups of speakers.⁴ In practice, both corpus studies and linguistic experimentation (e.g. careful elicitation of acceptability judgments) depend on pools of speakers to get statistically significant results (see recent practical discussions of the use of scientific methods in syntax in Cowart 1997 and Schütze 1996).

In studies of ongoing change and variation, one approach is to correlate acceptability judgments with the relative frequency of variants (Anttila 1997:12, Mohanani 2003). For example, consider a situation where a certain variant X is less frequent than another variant Y. If variant X is also less acceptable for an individual speaker than variant Y, then we might postulate that mental grammars are gradient, and that this gradience is transferred to performance.

This is not enough, though: 'we still have to factor out effects of other language use factors such as comprehension difficulties, direct effects of frequency on acceptability judgments, and so on' (Mohanani 2003). Boersma (2004) discusses mismatches between acceptability judgments and corpus frequencies, noting cases from phonology and syntax where corpus frequencies of a given variant are different from what acceptability judgments predict.⁵ He argues that these mismatches can be explained by differences

² The methodological obstacle that advocates of stochastic grammar face is reminiscent of Labov's (1994: 11) observation that '[h]istorical linguistics can be thought of as the art of making the best use out of bad data'. Thanks to Brian Joseph and Richard Oehrle for discussion of this issue.

³ This argument holds only if we assume speakers don't store information about who says what.

⁴ Richard Oehrle pointed out to me that an important industrial example of this is the (ongoing) development of robust broad-coverage speech-to-text systems such as Nuance, which depend entirely on multi-individual corpora.

⁵ Ultimately, we would like to find out whether or not we see differences between the frequencies of a given variant in spoken and written corpora and how these differences match up with acceptability judgments. It could be the case that acceptability judgments converge with spoken, but not written, corpus frequencies for a given syntactic variant. I overlook the crucial distinction between the modality of different corpora here and assume that, in principle, acceptability judgments can diverge from corpus frequencies regardless of the spoken or written nature of the corpus.

between the linguistic task of production and the task of providing acceptability judgments. For production, the speaker chooses a pronunciation that best harmonizes speaker- and listener-based requirements. By contrast, acceptability judgments involve choosing, for a given meaning, an overt form that comes closest to realizing it. Mismatches between acceptability judgments and production occur because of the role speaker-based preferences play in the production process.

Summing up, the force of N's methodological argument is weakened once we clarify what corpora can be employed for. N's argument depends on a typical property of present-day corpora: they are not tailored to particular individuals. This is an accidental property, not an essential one. Moreover, recent work suggests how one can use the frequency asymmetries in corpora to get at the nature of individual mental grammars, while avoiding the postulation of community grammars.

3. EXTRAGRAMMATICAL FACTORS DETERMINE STATISTICAL PREFERENCES. N's (2003: 696–97) second argument against stochastic grammar critiques the evidence for incorporating probabilistic information into grammar models. The evidence, he claims, is weak. Consider extraposition of sentential subjects, as in 1 (example from Manning 2003:307).⁶

- (1) a. It is unlikely that the company will be able to meet this year's revenue forecasts.
 b. #That the company will be able to meet this year's revenue forecasts is unlikely.

According to N, it is incorrect to say that because speakers are more likely to say 1a than 1b that this likelihood forms part of our mental grammar. Rather, speakers tend to avoid sentences like 1b because they are harder to process (Hawkins 1994).⁷

3.1. THE STOCHASTIC GENERALIZATION. For N, preferences like 1 reflect processing factors, rather than grammar-internal factors (Newmeyer 2003:696). Specifically, N is assuming a performance-based account like that discussed in Hawkins 1994. In this account, the preference in 1 is to be explained by pressure to shorten recognition time for phrasal constituents (Newmeyer 2003:684).⁸ But whether the factors that influence grammatical preferences are extragrammatical or not needs to be decided on a case-by-case basis. Many factors can influence the choice of one variant over another, for example, memory limitations, perception, articulation, and fatigue. However, it does not follow from the fact that many factors influence the choice of a syntactic variant that frequency of use of certain constructions is not represented directly in mental grammar.⁹

⁶ The '#' mark is used by Manning to indicate that people prefer to put *that*-clauses after the finite verb in clauses like 1.

⁷ Other factors apart from processing, such as real-world knowledge and context, play a role in explaining the preference for 1a. For example, Miller (2001) argues that there are discourse conditions on the choice between extraposition and nonextraposition. Nonextraposition (as in 1b) requires that the content of the subject be discourse-old or inferable. If the content of the subject is not discourse-old or inferable, then extraposition (as in 1a) is obligatory.

⁸ See also Wasow 2002:9–12, 25 for discussion of how extraposition is influenced by the weight of constituents and for different conclusions from N's about the implications for theories of mental grammar.

⁹ Thanks to Christopher Manning for discussion of this point.

Further, factors that affect the choice of variants in one language can often be found to have a categorical influence in other languages (see Givón 1979:26–31, Dik 1997: 34ff., and Bresnan et al. 2001). This observation has been called the STOCHASTIC GENERALIZATION (Bresnan & Aissen 2002b). Givón (1979:26) gives the example of a communicative tendency to reserve subject position for topics. In some languages (e.g. Krio) this tendency is expressed categorically (100%); in other languages (e.g. English) the same communicative tendency is expressed at a noncategorical level of 90% (Givón 1979:28).

The thrust of the argument from the stochastic generalization can be further illustrated by considering a candidate instance of the principle of end weight in English, discussed in Wasow 2002. The principle of end weight says that phrases are presented in order of increasing weight (Wasow 2002:3). This principle arguably holds for the canonical constituent order of English, as illustrated in 2 (slightly adapted from Gazdar & Pullum 1981:120 and cited in Wasow 2002:4). Since heads are single words, they can be lighter than phrases.¹⁰ Prepositional phrases can contain noun phrases as a proper subpart, so prepositional phrases can be heavier than noun phrases. Further, clauses and verb phrases can be heavier than prepositional phrases.

(2) $H < N'' < P'' < V''$

(Head < Noun phrase < Prepositional phrase < Verb phrase or clause)

Noncategorical phenomena like heavy NP-shift are exceptions to the canonical order in 2. In heavy NP-shift (e.g. *I loaned to Kim the movies that I rented*) a non-NP sits between a verb and a following heavy NP. Now recall that N assumes that in language use people are sensitive to quantitative information, but that mental grammars consist of only categorical information. This division of labor misses a generalization:

the Principle of End Weight, which depends on relative measures of length and complexity, manifests itself in paradigmatically grammatical phenomena: the canonical constituent order of English (and other languages) . . . If the canonical ordering and the obligatory cases are part of the competence grammar, but the quantitative preferences are treated as performance, then a larger generalization is lost. (Wasow 2002:139)

A similar argument is made in Givón 1979:26–31, Bresnan et al. 2001, and Bresnan & Nikitina 2003.

It is possible to argue, as N (1998:39–44) does, that there is no missed generalization: a single external factor exerts its influence on language users, which has categorical effects in some languages (where the results of the force are grammaticalized, that is, embedded in the grammar), but not in others (where speakers have multiple grammatical options).¹¹

This position can be characterized in the following way. Take two languages, L_1 and L_2 . In L_1 (e.g. English), both indefinite and definites are possible subjects. However, indefinite subjects are rare. In L_2 (e.g. Bemba (Givón 1979) and Malagasy (Keenan 1976)), indefinite subjects are prohibited. The unifying external factor that connects the distributions found in L_1 and L_2 is that there is a general communicative preference to put nominal expressions with contextually salient referents (e.g. definite NPs) in perceptually salient positions (subject or clause-initial position).

¹⁰ Number of words is only one measure of weight. Others are possible, for example, number of syllables (Green 2004). Under that metric, *elephantiasis* would be heavier than *eyes of newt*. Thanks to Brian Joseph for discussion of this point.

¹¹ Thanks to Fritz Newmeyer for discussing this position with me in personal communication.

As a counterpoint, consider the optimality-theoretic (OT) account of the interaction of definiteness and grammatical role in Aissen 2003. The constraint subhierarchy in 3 forms part of Aissen's account.

(3) *Su/NSpec >> *Su/Spec >> *Su/Def >> *Su/PN >> *Su/Pro

The two highest ranking constraints, *Su/NSpec and *Su/Spec, prohibit nonspecific indefinite and specific indefinite subjects, respectively.

The subhierarchy in 3 is grounded in cognitive and communicative tendencies and is claimed to be part of the synchronic mental grammars of speakers of both L_1 and L_2 . In L_1 , where the constraint subhierarchy is low-ranked, the effects of the constraints can be detected in only a restricted set of contexts (see Bresnan & Aissen 2002a:87–88 and the discussion below on the pronunciation of nonnative forms and pidgin genesis). In L_2 , the constraint subhierarchy is highly ranked and the constraints have a categorical effect (e.g. in Bemba and Malagasy).

Both N's account and Aissen's OT account postulate that there is a unifying explanation for the distribution of indefinite and definite NPs. In N's account external factors (communicative preferences) play the central role, whereas in Aissen's account an internal factor (a constraint subhierarchy) grounded in communicative preferences does.

I now turn briefly to some phenomena that favor the OT account over N's. There is a growing number of analyses of phenomena that appear to demonstrate that low-ranked and hence relatively inactive constraints are actually present in the synchronic state of individual mental linguistic systems (e.g. Bresnan 2000 and Davidson et al. 2004). For an account like N's, this type of phenomena would have to be explained by extragrammatical factors, for example, parsing ease, context, and real world knowledge and behavior. I discuss two phenomena—pronunciation of nonnative forms and pidgin genesis—that pose a challenge for N's account. For both of these phenomena, I make reference to language users reranking constraints. The relevant behavior for both of these cases could have occurred for various reasons, for example, accommodation through the elimination of marked structures that are inaccessible to interlocutors in the case of pidgin genesis (Thomason 1997, Thomason & Kaufman 1988:174ff.). This behavior is modeled in terms of the elevation of constraints. There is no sense in which the reranking of constraints CAUSES the relevant behavior (cf. McMahan 2000).¹²

The first case involves the pronunciation of nonnative forms. Davidson and colleagues (2004:337ff.) demonstrated that when faced with nonnative inputs, speakers can elevate faithfulness constraints from lower positions to higher positions. They asked adult English speakers to produce forms with initial clusters such as [kt] that are illegal in English. They found that for some clusters (e.g. [zr]), speakers were able to produce faithful, English-illegal outputs, with no nativization. On this basis, they proposed that when producing these English-illegal forms, speakers prioritize low-ranked faithfulness constraints above the relevant markedness constraints. This example suggests, contra N, that constraints that are ranked too low to have a categorical effect in a language are actually present in synchronic mental grammars.

The second case involves pidgin genesis. In situations where the contact languages are typologically distant (i.e. share few marked forms in common), creators of pidgins rerank low-ranked individual markedness constraints above conflicting faithfulness constraints (Bresnan 2000). Reranking the constraints in this way reflects the removal

¹² Thanks to Brian Joseph for raising this issue.

of marked forms (e.g. bound pronouns). Bresnan (2000) discusses markedness of pronoun forms. She assumes two types of constraints: first, a set of markedness constraints against complex or difficult structures, for example, constraints against pairing bound and zero forms with pronominal content; and second, a set of faithfulness constraints requiring that features of the input content be preserved in output forms so that, for example, reduced forms are specialized for topical content. In pidgin genesis, simplification is modeled as a process in which speakers eliminate marked features of their language by reranking low-ranked markedness constraints above individual faithfulness constraints that conflict with them. Constraints targeted for reranking are those that mark types of forms that are not understood or not easily learned by interlocuters because they are not in the inventory of the interlocuters' language.¹³ For example, if a speaker of a language in which bound forms realize topical, pronominal content is trying to communicate with a hearer that does not have bound forms in his/her language, then the speaker elevates the markedness constraint against bound forms paired with pronominal content.

In sum, N's argument assumes that factors that influence the choice of syntactic variants are extragrammatical. This runs afoul of the stochastic generalization—factors that affect the choice of variants have a categorical effect in other languages. Further, several recent analyses (Bresnan 2000, Davidson et al. 2004) suggest that low-ranked constraints can be shown to be synchronically active in certain contexts. This type of phenomenon poses a challenge to any approach that attributes the choice of variants solely to extragrammatical factors.

3.2. THE CONFLATION OF COMPETENCE AND PERFORMANCE. Newmeyer's argument that extragrammatical factors influence statistical preferences relates directly to another argument: stochastic grammars conflate competence and performance. According to this argument, in cases of optionality, stochastic grammars not only deliver the options, but also predict their frequency of occurrence. This assumption leads to the abandonment of the competence/performance distinction because, in order to predict actual frequencies, all the relevant performance factors would have to be put in the grammar model, that is, 'all constraints on perception, articulation, memory, fatigue, style, and politeness interact with grammatical constraints' (Keller & Asudeh 2002:240). For Keller and Asudeh (2002:240), this leads to absurd conclusions with respect to familiar claims of OT such as factorial typology because 'surely speakers with distinct native languages have cognitive abilities in common and these cannot be reranked to yield their different languages'.¹⁴ For N (2002:98), this assumption is a short step away from late generative semantics (Lakoff 1974), 'which argued that *all* factors relevant to morpheme occurrence in discourse are a matter for grammatical analysis'.

First, as pointed out above, it does not follow from the fact that lots of factors (including factors that should not be part of a system of constraints) influence the frequency of certain constructions that knowledge of the frequency of use of various constructions is not and should not be part of mental grammars.

Moreover, it is not true that a commitment to stochastic grammar forces one to abandon

¹³ In this way, Bresnan's model of simplification is hearer-based. See Thomason & Kaufman 1988:174ff. for extensive empirical support for the hypothesis that pidgin genesis is a result of mutual linguistic accommodation.

¹⁴ 'To construct a factorial typology of set of constraints, we sum up logically possible rankings of this set of constraints, and compute the different outcomes. With large sets of constraints the number of possible rankings rises steeply, as with a constraint set of size n , we must consider all $n!$ rankings. (This equals 2 rankings for 2 constraints, 6 rankings for 3, 24 for 4, 120 for 5, 720 for 6, etc.)' (Kager 1999:35)

the competence/performance distinction. I now give a brief overview of the framework of stochastic optimality theory (Boersma 1998, Boersma & Hayes 2001) and explain briefly how it preserves standard assumptions about competence and performance.

Stochastic optimality theory extends standard optimality theory (Prince & Smolensky 1993) in two ways. First, constraints are ranked on a continuous scale of real numbers, rather than a discrete ordinal scale. Thus, constraints not only dominate other constraints, but they are specific distances apart, and these distances are relevant to the predictions of the theory. Second, when candidates are evaluated to determine the optimal candidate, the rank of each constraint is slightly perturbed by adding to its rank a random value, called NOISE, from a normal distribution. For example, a constraint with the mean rank of 99 could be evaluated at 98.1 or 100.3. It is the constraint ranking that results from these sampled values that is used in evaluation.

In stochastic optimality theory, constraints act as if they were associated with a range of possible values rather than single points. The value that is used at evaluation time is called an EVALUATION POINT. The rank more permanently associated with a constraint (i.e. the center of the range) is called a RANKING VALUE.

Figure 1 illustrates two constraints, C_1 and C_2 . The ranking values of the two constraints are the means of their varying rankings and are marked at the top of the two bell curves; thus, normally, $C_1 \gg C_2$. On some evaluations, however, the evaluation point of C_1 will fall in the lower end of its normal distribution at the same time that the evaluation point of C_2 falls at the higher end of its distribution, giving the effect of reranking as C_2 has precedence there over C_1 , with $C_2 \gg C_1$. Over a sequence of evaluations, overlapping ranges have a significant effect: for inputs in which $C_1 \gg C_2$ C_2 produces a different output than $C_2 \gg C_1$, multiple outputs for a single input are observed. In this way, stochastic optimality theory models intraspeaker variation.

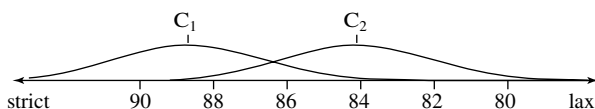


FIGURE 1. Constraint ranking on a continuous scale with stochastic evaluation.

Stochastic optimality theory preserves the competence/performance distinction in the following way (Bresnan & Aissen 2002b). Each grammar model (mental grammar represented as a set of constraints and their (mean) ranking values) is embedded in a USAGE GRAMMAR. I define a usage grammar as consisting of a grammar model plus some mechanism that allows extralinguistic factors (e.g. politeness, exhaustion) to affect the ranking values of constraints. For some of these factors (e.g. style), a systematic mechanism may be needed to link these factors to particular constraints (see below). For other nonsystematic factors (e.g. exhaustion), we need a model of our ignorance of the entire context and the nonlinguistic factors that condition the probability of an output. The noise variable of stochastic evaluation may be taken as a model for these nonsystematic performance effects.¹⁵ Extragrammatical factors are represented in the usage grammar by variables that perturb the rankings of constraints at evaluation time, NOT as individual constraints.

¹⁵ However, to conclude that the intraspeaker variability modeled by stochastic grammars is random and uncaused is 'the fallacy of reified ignorance' (Bresnan & Deo 2001:37): 'stochastic models represent gaps in our knowledge of the world, not gaps in the causal structure of the world'. Thanks to Joan Bresnan for discussion of this issue.

4. THE MULTIPLICITY OF GENRES. N's third argument starts from the observation that the probability of use of a syntactic variant is a function of a particular genre or style (Newmeyer 2003:697–98). N points out, however, that there are a large number of genres (Biber 1995). Proponents of stochastic grammar are forced to claim that the mental grammars of speakers who command certain genres differ from those of others who do not. Rather, concludes N, we should distinguish knowledge and (appropriate) use, which stochastic grammars fail to do.

The particular example that N (2003:697–98) uses is the 'bizarre (but genre-normal) syntax and lexicon' from flight crews. Example 4 gives some of N's examples.

- (4) a. We are ready to depart the gate.
 b. Takeoff will be shortly.
 c. We hope that you will enjoy your stay in the Seattle area or wherever your destination may take you.

If the examples in 4 are part of a small set of memorized formulae, then they are not relevant to stochastic grammar. Rather, all we need to say is that people learn a set of such formulae during their experience in the airline industry. Possible support for this claim would come from demonstrating that examples similar to 4 show up across different airlines. However, if the examples can be demonstrated to be a reflection of structured variation, then the issue becomes how to model stylistic and genre variation in stochastic grammar.

N's argument assumes that stochastic grammars (e.g. via stochastic evaluation, probability distributions over argument frames, etc.) completely derive output distributions. This is problematic, states N, as the information encoded in these grammar models might be epiphenomenal, because extragrammatical factors, like genre and style, actually determine the probability of use of a particular syntactic variant. It is this last point that reveals another underlying assumption of N's argument: mental grammars do not include a mechanism that systematically links extralinguistic factors to linguistic variables.

As a starting hypothesis, let's assume that there is no systematic relationship between grammatical constraints and extragrammatical factors like genre and style. Exactly this is reflected in the basic architecture of stochastic optimality theory, as summarized in §3. In stochastic optimality theory, mental grammar is modeled as the ranking values of individual constraints. Specific occurrences of linguistic variants are modeled as random and unpredictable. These performance effects are modeled by the normal distribution used in the stochastic evaluation of the constraints, as in Fig. 1.

The evaluation point (*SELECTIONPOINT*) of a constraint C_i at evaluation time is given by the equation in 5 (adapted from Boersma 2000:483).

$$(5) \textit{selectionPoint}_i = \textit{rankingValue}_i + \textit{NOISE}$$

There is good reason to believe that this picture is overly simplistic. Weinreich and colleagues (1968:169) define a strict condition under which a linguistic variable can be defined: 'Quantitative evidence for covariation between the variable in question and some other linguistic or extralinguistic element provides a necessary condition for admitting such a structural unit'. One of the central observations of the variationist tradition is that knowledge of language includes sensitivity to linguistic variation that correlates with different styles and genres, APPROPRIATELY DEFINED.¹⁶

¹⁶ Recent work has problematized the classical notion (Labov 1966) of reified style categories; see the papers in Eckert & Rickford 2001.

Variable rules (Labov 1969, Paolillo 2002) are one type of mechanism that links extralinguistic factors to linguistic variables. Covariation between linguistic variables and extralinguistic elements, like style and genre, can be modeled in stochastic optimality theory by allowing these elements to systematically boost or depress the ranking values of selected constraints. Boersma and Hayes (2001:82–83) discuss how this might be implemented.¹⁷

We assume that utterances occur in contexts that can be characterized along a casual-to-formal continuum. We quantify this continuum with a variable *Style*, such that *Style* equals 0 in maximally casual speech and 1 in maximally formal speech. The selection point for a given constraint C_i is determined by [the following equation:]

$$\textit{selectionPoint}_i = \textit{rankingValue}_i + \textit{styleSensitivity}_i \times \textit{Style} + \textit{NOISE}$$

This is the same equation as before [example 5 above], augmented by the term $\textit{styleSensitivity}_i \times \textit{Style}$, in which $\textit{styleSensitivity}_i$ is a constraint-specific value. Constraints with positive values for $\textit{styleSensitivity}_i$ take on higher ranking values in formal speech; constraints with negative values for $\textit{styleSensitivity}_i$ take on higher ranking values in casual speech, and constraints with zero values of $\textit{styleSensitivity}_i$ are style insensitive.

Note that the **STYLESENSITIVITY** factor is not itself a constraint. Rather, it systematically boosts or depresses the ranking of specific constraints or groups of constraints. In this way, it makes good on the following ‘general statement’ by Weinreich and colleagues (1968:187–88):

Linguistic structure includes the orderly differentiation of speakers and styles through rules which govern variation in the speech community; NATIVE COMMAND OF THE LANGUAGE INCLUDES THE CONTROL OF SUCH HETEROGENEOUS STRUCTURES. [emphasis mine]

Given the model that Boersma and Hayes lay out in the quote above, speakers and hearers who command certain styles and genres do not have a different constraint set than speakers and hearers who do not, contra N. Rather, systematic covariation between linguistic variables and extralinguistic factors is modeled in terms of weights that alter the ranking of constraints.

In sum, stochastic optimality theory without a ‘styleSensitivity’ factor treats the output distribution of linguistic variables as a performance effect, as advocated for by N. However, one of the major insights of the variationist tradition is that linguistic variables covary systematically with extralinguistic factors like style and genre. The addition of the ‘styleSensitivity’ factor to the basic architecture of stochastic optimality theory is one way of making good on this insight.

5. CONCLUSION. I have examined three arguments against stochastic grammar from Newmeyer 2003. The overarching conclusion is that none of these arguments are sufficient to refute the guiding assumptions of the inherent variability tradition: knowledge of language accommodates and generates variation, and mental grammar includes a quantitative, noncategorical, and nondeterministic component.

REFERENCES

- ABNEY, STEVEN. 1996. Statistical methods and linguistics. *The balancing act: Combining symbolic and statistical approaches to language*, ed. by Judith Klavans and Philip Resnik, 1–26. Cambridge, MA: MIT Press.
- AISSEN, JUDITH. 2003. Differential object marking: Iconicity vs. economy. *Natural Language and Linguistic Theory* 21.435–83.

¹⁷ See Lee 2002 for discussion of how Boersma and Hayes’s (2001) stochastic optimality theory model of stylistic variation can be applied to intraspeaker variation in the ellipsis of case markers in Korean.

- ANTTILA, ARTO. 1997. *Variation in Finnish phonology and morphology*. Stanford, CA: Stanford University dissertation.
- BENDER, EMILY M. 2001. *Syntactic variation and linguistic competence: The case of AAVE copula absence*. Stanford, CA: Stanford University dissertation.
- BIBER, DOUGLAS. 1995. *Dimensions of register variation: A cross-linguistic comparison*. Cambridge: Cambridge University Press.
- BOERSMA, PAUL. 1998. *Functional phonology: Formalizing the interactions between articulatory and perceptual drives*. The Hague: Holland Academic Graphics.
- BOERSMA, PAUL. 2000. Learning a grammar in functional phonology. *Optimality theory: Phonology, syntax, and acquisition*, ed. by Joost Dekkers, Frank van der Leeuw, and Jeroen van de Weijer, 465–523. Oxford: Oxford University Press.
- BOERSMA, PAUL. 2004. A stochastic OT account of paralinguistic tasks such as grammaticality and prototypicality judgments. Amsterdam: University of Amsterdam, ms.
- BOERSMA, PAUL, and BRUCE HAYES. 2001. Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32.45–86.
- BRESNAN, JOAN. 2000. Pidgin genesis and optimality theory. *Processes of language contact: Case studies from Australia and the Pacific*, ed. by Jeff Siegel, 145–73. Montreal: Les éditions Fides.
- BRESNAN, JOAN, and JUDITH AISSÉN. 2002a. Optimality and functionality: Objections and refutations. *Natural Language and Linguistic Theory* 20.81–95.
- BRESNAN, JOAN, and JUDITH AISSÉN. 2002b. Optimality theory and typology. (Course notes from summer school on formal and functional linguistics.) Duesseldorf: University of Duesseldorf, ms.
- BRESNAN, JOAN, and ASHWINI DEO. 2001. Grammatical constraints on variation: 'Be' in the survey of English dialects and (stochastic) optimality theory. Stanford, CA: Stanford University, ms.
- BRESNAN, JOAN; SHIPRA DINGARE; and CHRISTOPHER D. MANNING. 2001. Soft constraints mirror hard constraints: Voice and person in English and Lummi. *Proceedings of the LFG01 Conference*, ed. by Miriam Butt and Tracy H. King, 13–32. Stanford, CA: CSLI Publications.
- BRESNAN, JOAN, and TATIANA NIKITINA. 2003. On the gradience of the dative alternation. Stanford, CA: Stanford University, ms.
- BREW, CHRIS, and MARC MOENS. 2000. Data-intensive linguistics. Edinburgh: University of Edinburgh, ms.
- CLARK, BRADY Z. 2004. *A stochastic optimality theory approach to syntactic change*. Stanford, CA: Stanford University dissertation.
- COWART, WAYNE. 1997. *Experimental syntax: Applying objective methods to sentence judgments*. Thousand Oaks, CA: Sage Publications.
- DAVIDSON, LISA; PETER JUSCZYK; and PAUL SMOLENSKY. 2004. The initial and final states: Theoretical implications and experimental explorations of richness of the base. *Constraints in phonological acquisition*, ed. by René Kager, Joe Pater, and Wim Zonneveld, 321–68. Cambridge: Cambridge University Press.
- DICK, FREDERIC, and JEFFREY L. ELMAN. 2001. The frequency of major sentence types over discourse levels: A discourse analysis. *Newsletter of the Center for Research in Language, University of California, San Diego* 13.3–19.
- DIK, SIMON C. 1997. *The theory of functional grammar. Part 1: The structure of the clause*. Berlin: Mouton de Gruyter.
- ECKERT, PENELOPE, and JOHN R. RICKFORD (eds.) 2001. *Style and sociolinguistic variation*. Cambridge: Cambridge University Press.
- ERNESTUS, MIRJAM, and R. HARALD BAAYEN. 2003. Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language* 79.5–38.
- GAZDAR, GERALD, and GEOFFREY PULLUM. 1981. Subcategorization, constituent order and the notion 'Head'. *The scope of linguistic rules*, ed. by Michael Moortgat, Harry van der Hulst, and Teun Hoekstra, 107–223. Dordrecht: Foris Publications.
- GIVÓN, TALMY. 1979. *On understanding grammar*. New York: Academic Press.
- GREEN, GEORGIA M. 2004. Review of *Postverbal behavior*, by Thomas Wasow. *Language* 80.327–31.
- HAWKINS, JOHN A. 1994. *A performance theory of order and constituency*. Cambridge: Cambridge University Press.

- KAGER, RENÉ. 1999. *Optimality theory*. Cambridge: Cambridge University Press.
- KEENAN, EDWARD. 1976. Towards a universal definition of 'subject'. *Subject and topic*, ed. by Charles Li, 303–33. New York: Academic Press.
- KELLER, FRANK, and ASH ASUDEH. 2002. Probabilistic learning algorithms and optimality theory. *Linguistic Inquiry* 33.225–44.
- LABOV, WILLIAM. 1966. *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.
- LABOV, WILLIAM. 1969. Contraction, deletion, and inherent variability of the English copula. *Language* 45.715–62.
- LABOV, WILLIAM. 1994. *Principles of linguistic change: Internal factors*. Cambridge, MA: Blackwell.
- LAKOFF, GEORGE. 1974. Interview. *Discussing language: Dialogues with Wallace L. Chafe, Noam Chomsky, Algirdas J. Greimas [and others]*, ed. by Herman Parret, 151–78. The Hague: Mouton.
- LEE, HANJUNG. 2002. Referential accessibility and stylistic variation in OT: A corpus study. *Chicago Linguistic Society* 38.361–78.
- MANNING, CHRISTOPHER D. 2003. Probabilistic syntax. *Probabilistic linguistics*, ed. by Rens Bod, Jennifer Hay, and Stefanie Jannedy, 289–341. Cambridge, MA: MIT Press.
- MANNING, CHRISTOPHER D., and HINRICH SCHÜTZE. 1999. *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.
- MCMAHON, APRIL. 2000. *Change, chance, and optimality*. Oxford: Oxford University Press.
- MILLER, PHILIP H. 2001. Discourse constraints on (non)extraposition from subject in English. *Linguistics* 39.683–701.
- MOHANAN, K. P. 2003. Objects of inquiry, theories, data, and methodology in linguistics. Stanford, CA: Stanford University, ms.
- NAGY, NAOMI, and BILL REYNOLDS. 1997. Optimality theory and variable word-final deletion in Faetar. *Language Variation and Change* 9.37–55.
- NEUMEYER, FREDERICK. 1998. *Language form and language function*. Cambridge, MA: MIT Press.
- NEUMEYER, FREDERICK. 2002. A rejoinder to Bresnan and Aissen. *Natural Language and Linguistic Theory* 20.97–99.
- NEUMEYER, FREDERICK. 2003. Grammar is grammar and usage is usage. *Language* 79.682–707.
- PAOLILLO, JOHN C. 2002. *Analyzing linguistic variation*. Stanford, CA: CSLI Publications.
- PRINCE, ALAN, and PAUL SMOLENSKY. 1993. Optimality theory: Constraint interaction in generative grammar. (Report no. RuCCS-TR-2.) New Brunswick, NJ: Rutgers University Center for Cognitive Science, ms.
- SCHÜTZE, CARSON T. 1996. *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago: The University of Chicago Press.
- SKOUSEN, ROYAL. 1989. *Analogical modeling of language*. Dordrecht: Kluwer.
- THOMASON, SARAH G. 1997. A typology of contact languages. *The structure and status of pidgins and creoles*, ed. by Arthur K. Spears and Donald Winford, 71–88. Amsterdam: John Benjamins.
- THOMASON, SARAH G., and TERRENCE KAUFMAN. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- WASOW, THOMAS. 2002. *Postverbal behavior*. Stanford, CA: CSLI Publications.
- WEINREICH, URIEL; WILLIAM LABOV; and MARVIN HERZOG. 1968. Empirical foundations for a theory of language change. *Directions for historical linguistics: A symposium*, ed. by Winfred P. Lehmann and Yakov Malkiel, 95–188. Austin: University of Texas Press.
- YANG, CHARLES D. 2003. *Knowledge and learning in natural language*. Oxford: Oxford University Press.

Northwestern University
2016 Sheridan Rd.
Evanston, IL 60208-2150
[bzack@northwestern.edu]

[Received 14 October 2003;
revision invited 15 December 2003;
revision received 3 May 2004;
accepted 19 September 2004]