to appear in *Lingua*, 2007

Innateness, Universal Grammar, and Emergentism*

William O'Grady

1. Introduction

Controversy over the nature of the innate endowment for language is an ongoing theme in the literature on language acquisition. While it might appear that this controversy reflects the existence of fundamental unresolved issues in the field, the areas of disagreement are more restricted than initial appearances suggest.

In point of fact, there is general agreement that the acquisition of language is innately guided—this much has been widely acknowledged even by those opposed to the idea of an innate Universal Grammar (e.g., Slobin 1985:1158, Elman et al. 1996:31 & 41, MacWhnney 2002, Tomasello 2003: 40 & 283). Absolutely no one claims that language can be acquired by anything other than a human brain, and there is no disputing the fact that congenital neurological pathologies of various sorts can impede the acquisition and use of language.

There is then essential agreement on Chomsky's (1975:13) observation that 'every "theory of learning" that is even worth considering incorporates an innateness hypothesis.' The debate is over the formulation of that innateness hypothesis—disagreements over the inborn endowment for language turn on its character, not its existence. Let us begin by considering a relatively uncontroversial case.

The most basic appeal to innateness involves the need for a 'vocabulary' of concepts and contrasts in terms of which generalizations can be formulated. Gregg (2003:120ff) makes the point with great effectiveness in his sweeping critique of emergentist approaches to second language acquisition, and the observation is common in the literature on first language acquisition as well. As Bowerman (1987:461) succinctly puts it, 'how does the child identify [grammatically relevant] features as important if they are not present ahead of time?' Jusczyk (1997:198) makes a parallel point with regard to speech perception:

> Out of all the possible ways in which learners could conceivably categorize the speech signal, some dimensions are favored, and others are not… [Infants] are counting the right sorts of properties in any distributional analysis that they perform on the input…

Given that, minimally, language involves a relationship between phonetic forms on the one hand and semantic representations on the other, we can be sure that

* I thank Kevin Gregg for helpful comments, Robert Bley-Vroman for advice on the corpus data, and Elwood Mott for sharing his expertise on Chinese.

generalizations about its workings will have to refer both to phonetic notions (alveolar, voiced, vowel, etc.) and to semantic concepts (plural, definite, stage-level, telic, and so forth). Moreover, it is hard to imagine an account of the availability of such concepts and contrasts that would not assume some sort of innate perceptual system and some sort of innate conceptual system. And indeed a neurobiological contribution in these areas is very widely accepted, if not always stated. What is open to dispute, however, is the question of whether and to what extent the innate components of the perceptual and conceptual systems qualify as instantiations of 'Universal Grammar.'

Take the perceptual system, for instance, and the work of Jusczyk (1997), who documents infants' aptitude for speech perception in great detail, making a very strong case for what he calls 'innately guided learning.' Yet, when considering the question of whether that aptitude is specific to language, he notes that the evidence does not point unequivocally toward the existence of specialized mechanisms for *speech* (p. 64). Indeed, there is considerable evidence that much of the perceptual system used for speech is present in other primates (e.g., Kuhl & Miller 1975, Raums et al. 2000, Hauser et al. 2001), leading Fitch, Hauser & Chomsky (2005:195-96) to conclude that:

> The safest assumption, at present, is that the mechanisms underlying human speech perception were largely in place before language evolved, based on either general auditory or vocalization-specific perceptual processes.

A parallel point can be made concerning the conceptual system. Language surely makes use of innately structured semantic representations, but it is far from clear that the effects of *Universal Grammar* can be discerned here. For one thing, there is the possibility, advocated by Jackendoff (2002:123), that semantic representations are not even *linguistic* objects.

> Conceptual structure is not part of language per se—it is part of thought. It is the locus for the understanding of linguistic utterances in context, incorporating pragmatic considerations and "world knowledge"; it is the cognitive structure in terms of which reasoning and planning take place.

Moreover, as Fitch, Hauser & Chomsky (2005:191) note, there is a very real possibility that 'the conceptual structure expressed by language is based upon a foundation shared with other animals'—a point with which Pinker & Jackendoff (2005:206) agree. 'The overall picture,' Pinker & Jackendoff suggest, is that:

> there is a substrate of conceptual structure in chimps, overlain by some uniquely human but not necessarily language-based subsystems, in turn overlain by subsystems that depend on the pre-existence of linguistic

> expression. [e.g., number, kinship, intervals of time (such as 'week'), and social roles such as 'justice of the peace' and 'treasurer']

A similar view is has been put forward within developmental psychology by Hespos & Spelke (2004:455):

> … the early development of semantic categories parallels the development of phonological categories and suggests that natural language semantics, like natural language phonology, evolved so at to capitalize on pre-existing representational capacities.

In sum, it seems unlikely that the case for *grammatical* nativism would ever have been made (or so widely accepted) if it had been based solely on the argument that notions such as definiteness, specificity, person, and number are in some sense innately given or that contrasts involving voice onset time, vowel height, and pitch reflect deep-seated properties of the biology of perception. If nothing more than this was claimed, the term 'Universal Grammar' would never have been proposed and the controversy that has dominated work in acquisition theory would not have arisen.

The *sine qua non* of Universal Grammar is not a claim about the need for a vocabulary of semantic concepts and phonetic contrasts; it is a claim about the nature of the computational system that structures form and relates it to meaning. In particular, it is the claim that this system includes faculty-specific principles that cannot be learned from experience. Classic examples of such principles include the binding principles (e.g., an anaphor must be bound in its governing category), constraints on contraction (e.g., a Case-marked trace blocks contraction), scope principles (e.g., a quantifier's scope corresponds to its c-command domain), conditions on displacement (e.g., the Subjacency Condition and the Empty Category Principle), and the like.

It makes sense to focus the 'innateness debate' on this point, for it is here that the strong claim has been made and it is here (and only here) that it clearly makes sense to talk about Universal Grammar in anything like the original sense of an innate system of grammatical principles. If the term 'Universal Grammar' comes to mean simply 'the innate endowment for language' or comes to be equated with any version of the 'innateness hypothesis' without regard for its content or character, its central thesis becomes a virtual truism—surely not the original intent of the nativist program for language.

Ironically, it is no longer clear that Chomsky believes that Universal Grammar in its original strong sense exists. As I noted in my opening paper, he has recently suggested that the properties of language might be reducible 'to properties of the interface system and general considerations of computational efficiency and the like' (2005:10). Nonetheless, it is clear that many other proponents of the Principles-and-Parameters framework continue to hold that UG as a *grammatical*

system does exist. Recent UG-based work includes countless proposals about grammatical principles of various sorts—see, for example, Baltin's (2006) analysis of VP preposing or Matushansky's (2006) proposal concerning the properties of head movement and its relationship to the Phase Impenetrability Condition. Moreover, leading figures in the field of acquisition research (e.g., Crain, Goro, & Thornton 2006:31) continue to hold that 'children are born with a set of universal linguistic principles' and that 'many aspects of adult grammar are innate.'

2. The case for UG

Not surprisingly, the UG-inspired contributions to this issue of *Lingua* raise a number of points in favor of the classic UG position and against the emergentist approach. Some of these points need to be set aside, for the reasons already noted, since they are actually arguments in favor of a general innateness hypothesis, rather than UG per se. However, many other points are directly relevant to the debate over the necessity of UG and the viability of emergentism—far too many to permit a thorough discussion here. I will briefly focus on three such points, each of which also relates to the role that frequency considerations might have in an emergentist account of how language works and how it is acquired.

2.1 Frequency and the processor—*want-to* contraction again

As mentioned in the introductory article, frequency is one of the factors to which emergentists look for their explanations. Although not inherently objectionable (everyone agrees that there are frequency effects in language), there is very substantial disagreement within linguistics over the importance of frequency to our overall understanding of how language works and how it is acquired.

All three of the pro-UG contributions to this special issue express doubts about the general explanatory potential of frequency, denying its relevance to the particular phenomena that they consider—consistent with a long-standing aversion within UG-related work to explanations of this sort.

Methodologically, appeals to frequency within emergentism are no different than appeals to government within Principles-and-Parameters theory. There is nothing inherently wrong with either idea. Each must stand or fall on its merits, and neither constitutes the foundation of the larger framework of claims and assumptions within which it is embedded. An argument against a significant role for frequency in language acquisition is not an argument against emergentism per se, anymore than an argument against government is an argument against UG per se. It is simply a consideration that needs to be taken into account in assessing the relevance of that factor to whatever theory of the language faculty one favors.

As I see it, frequency does in fact have a very important role to play in understanding how language works—but only if we acknowledge that its effects are modulated by the efficiency-driven processor that is at the heart of the language

faculty. The syntax of *want-to* contraction, discussed in detail in the article by O'Grady, Nakamura & Ito in this issue, is a case in point.

Frequency clearly has an important role to play in understanding why particular pairs of elements contract and others do not (e.g., Bybee & Schiebman 1999:576, Bybee 2002:112). It cannot be a coincidence that, as Krug (1998:294) observes, almost all of the full forms currently involved in contraction in English come from among the 150 most frequent words in the language. And, in fact, frequency may well have a small but significant role to play in understanding *want-to* contraction as well. Take, for example, the fact that *want to* contracts, but the homophonous *wont to* does not.

(1) a. I want to (wanna) go to the party.
b. I am wont to (*wanna) go to the party.

A search of the CHILDES data base for Adam, Eve and Sarah reveals an unsurprising, and presumably not unrelated, asymmetry in relative frequency: there are 980 instances of *want to,* but none of *wont to.*

But can frequency explain how and why speakers of English come to permit *want-to* contraction in certain contexts, but not others—as illustrated by the familiar contrast in (2)?

(2) a. Who do you want to (wanna) see?
b. Who do you want to (*wanna) stay?

Rohde (1999) suggests that it can, claiming that even small differences in frequency permit induction of contrasts between patterns. As a demonstration of this, he trained a Simple Recurrent Network on a one-million sentence corpus in which the various *want-to* patterns were represented in the following proportions. (These proportions were based on the relative rate of occurrence of the different patterns in the CHILDES data base.)

(3)a. <N TO> pattern; no contraction:
Who do you want to play with? [0.0012]

b. <N TO> pattern; with contraction:
Who do you wanna play with? [0.0008]

c. <N N TO> pattern; no contraction:
Who do you want to play with you? [0.0002]

d. <N N TO> pattern; with contraction:
Who do you wanna play with you? [0.0]

Rohde reports (pp. 18-19) that the SRN was significantly less likely to produce the (d) pattern than the other three patterns, suggesting that it was sensitive to the prohibition against contraction in this case.

I am skeptical about the significance of this result for a number of reasons. For one thing, it is not clear that everyone has access to even the relatively minimal amounts of experience that Rohde supposes. Bailes (2000) makes a particularly interesting observation in this regard, noting that speakers of British English, who are apparently not exposed to relevant data at all (British English does not permit any type of *want to* contraction), can nonetheless distinguish the 'good' cases from the 'bad' ones.

It is also unclear whether the SRN used by Rohde can generalize to cases where the *wh* word, the contraction site, and the gap are not in such close proximity as in the sorts of sentences found in his input.

(4) a. Prototypical pattern for prohibited contraction: the *wh* word is in the same clause as the potential contraction site.
Who did they **want to** stay at their home?

b. Departure from the prototype: the *wh* word is in a higher clause than the potential contraction site.
[**Who** did the host family say [that they **want to** stay at their home]]?

(5)a. Prototypical pattern for permissible contraction: the *wh* word is associated with the verb that serves as complement to *want*..
Who did the visitor want to **see**?

b. Departure from the prototype: the *wh* word is associated with a more deeply embedded verb.
Who do they want [to believe that [the visitor **saw**]]?

An even more serious problem is that the Rohde's SRN model offers no explanation for why the contrast in contractability exists in the first place. If it had been trained on input in which contraction occurred some of the time in the <N N to> pattern but never in the <N TO> pattern (the reverse of what happens in actual English), it would have no doubt learned the contrast illustrated in (6).

(6) 'Backward' English:

a. The <N N TO> pattern—contraction permitted.
Who do you wanna stay?

b. The <N TO> pattern—no contraction.
Who do you want to see?

Yet, no speaker of English talks this way, and it is hard to believe that any speaker of English ever could. In order to explain this fact, we need to refer to something other than frequency; we need an explanation for why some things are infrequent (and even non-existent) in the first place. If I'm right, the explanation lies in the interaction of an efficiency-driven linear processor with a physiologically motivated constraint on phonological reduction: contraction is most natural where the elements involved combine immediately.

This points to an intriguing, but highly plausible, state of affairs. If children are equipped with the type of processor and articulatory physiology that I assume, they are not dependent on Universal Grammar to acquire the contrast that we have been considering. Nor are they dependent on subtleties of the input. Those subtleties may well be there, at least some of the time, and it may well be possible to design computer-based networks that will record and respond to them. Nonetheless, on the view I propose, children adopt a more direct and more secure approach: a predisposition to treat contraction in a particular way emerges from the general cognitive and physiological resources with which they are endowed. That is how they come to make the particular contrast in contractability that they do. And, equally importantly, that is why the contrast comes to be present (however sparely) in the input in the first place.

2.2 An agreement puzzle

In his commentary, Hawkins (this issue) raises a curious fact about agreement and contraction in English—many speakers use the contracted form *there's*, but not its uncontracted counterpart, with a plural NP.[1]

(7) a. There's two books on the table.
b. *There is two books on the table.

A comparable phenomenon is also possible with *here's*.

(8) a. Here's three articles that you should read.
b. *Here is three articles that you should read.

This phenomenon exhibits at least two key properties.

First, contraction without agreement is possible only with singular forms—the plural forms *there're* and *here're* exhibit the expected plural agreement dependency.

[1]There are dialects in which patterns such as (7b) are acceptable—Ozark English (Wolfram & Christian 1976:78 & 82), for instance, but they are not relevant here. See O'Grady (to appear) for some discussion.

(9) *There're a book on the table.
(cf. There're books on the table.)

(10) *Here're an article that you should read.
(cf. Here're two articles that you should read.)

Second, contraction without agreement manifests lexical idiosyncrasies. Although fully natural with *there's* and *here's*, it sounds noticeably less acceptable, at least to my ear, with *where's* and seems to be entirely impossible with *when's*.

(11) ?*Where's the three articles that I'm supposed to read?
(cf. Where's the article that I'm supposed to read?)

(12) *When's the two concerts that you want to go to?
(cf. When's the concert that you want to go to?)

In sum, it appears that *there's* and *here's* (and to a lesser extent *where's*) show an 'anti-agreement effect'—they manifest the default third person singular form regardless of the person and number properties of the accompanying NP. How can we account for this?

One possibility is that *there's* and *here's* are showing the effects not just of contraction but also of grammaticalization, a process that often accompanies contraction and is also highly sensitive to frequency (e.g., Bybee 2003, Tomasello 2003:15). An obvious example of the combined effect of the two processes can be seen in English *gonna* (< *going to*), which has become a marker of the immediate future.

Another precedent, more directly relevant to the case at hand, involves French *voilà* (lit. 'see there') and *voici* (lit. 'see here'), which resemble their English counterparts *there's* and *here's* in not exhibiting agreement (Morin 1985). Instead, they function as invariant functors that combine with an NP to yield a clause.

(13)a. With a singular NP:
Voici/voilà un livre.
here.is/there.is a book
'Here's/there's a book.'

b. With a plural NP:
Voici/voilà deux livres.
here.is/there.is two books
'Here's/there's two books.'

If this idea is on the right track and if the prospects for grammaticalization, like those of contraction, are enhanced by frequency of usage, as we have been assuming, a number of predictions follow.

a. The forms *there's* and *here's*, which show the anti-agreement effect that we attribute to grammaticalization, should be more frequent than their respective plural counterparts, which show no such effect. This is correct. As the following summary of a Google search of the English-language Internet in late January of 2007 shows, *there's* and *here's* are a few hundred times more frequent than *there're* and *here're*, respectively.[2]

there's	300,000,000
there're	1, 320,000
here's	289,000,000
here're	525,000

b. *There's and here's*, which show the anti-agreement effect, should be more frequent than *where's*, which shows the effect only weakly. This too is correct—*where's* occurs approximately a third less often than *there's* and *here's*.

there's	300,000,000
here's	289,000,000
where's	195,000,000

c. *When's*, which does not show the effect at all, should in turn be significantly less frequent than *where's*. This too is correct: *when's* occur far less often than *where's*.

where's	195,000,000
when's	2,270,000

Of course, there is much more to the contraction and grammaticalization than frequency. To begin, one must ask why and how contractions such as *there's*, *here's*, *where's*, and *when's* come about in the first place. This in itself is not a trivial question since, as Bybee (2002:124) observes, it appears that the pairs of elements that undergo fusion in these cases (an expletive subject plus a verb, a locative plus a verb, and a *wh* word plus a verb) do not even form a constituent—a

[2]There is no comparable difference for the uncontracted forms--*there is* and *there are* each occur about a billion and a half times, for instance. It seems plausible to suppose that phonological considerations discourage contraction of the latter pair of elements, whose reduced form contains an [r] immediately followed by a syllabic [r].

rather counterintuitive state of affairs. In fact, though, Bybee may be wrong about this.

In the sort of linear computational system that I propose, the first two words in sentences such as *There's ink on the floor* and *Where's Mary?* do in fact enter into a fleeting combinatorial relation with each other as the sentence is being built.

(14)a.	Combination of *there* and *is*: [There is]	Combination of *where* and *is*: [Where is]
b.	Combination of *is* and *ink*: [There [is ink]]	Combination of *is* and *Mary*: [Where [is Mary]]

Because *there/where* and *is* combine immediately during the first step in the sentence formation process, creating a temporary constituent, contraction at this point satisfies the naturalness constraint proposed by O'Grady, Nakamura & Ito (this issue): contraction of the string XY is natural when X combines with Y without delay. Frequency enters into play only secondarily in that the relevant frequency of these particular computational routines may well have paved the way for eventual grammaticalization, which in turn results in the anti-agreement effect noted by Hawkins.

2.3 Functional morphology in SLA

In their contribution to this issue, Goad & White (see also Goad & White 2006) investigate the ability of native Mandarin speakers to produce three types of past tense forms in English.

i) irregular pseudo-inflection: *kept*
ii) regular inflection with a short stem: *wrap+ed*
iii) regular inflection with a long stem: *help+ed*

They report two key findings:

a. The most common pattern is for past tense inflection to be realized with a fortis release (hence [kɛpt^{h}], [ræp-t^{h}], etc.).

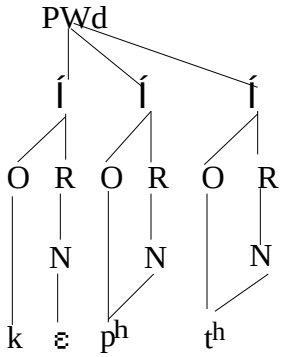
b. The final consonant of the stem frequently manifests a fortis release in regularly inflected forms, but not in pseudo-inflected forms (hence [ræp^{h} -t^{h}], but not *[kɛp^{h}t^{h}]).

Goad & White note that in the prosodic representations that they posit, the unattested form has an unlicensed empty nucleus, in violation of the Empty Category Principle—a component of Universal Grammar.

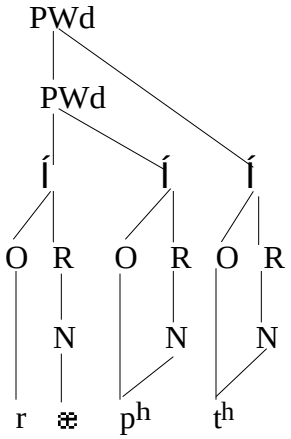
According to this theory, an empty nucleus must be followed by a domain boundary. This happens in (15b), Goad & White's representation for *wrapped*, in which each empty nucleus appears immediately before a prosodic word boundary. However, the principle fails to be satisfied in (15a), their representation for *kept*, in which both empty nuclei fall within the same prosodic word domain.

(15)

a.



b.



Goad & White take this result to constitute a challenge to emergentism, claiming that the stem-final consonants would all 'be analysed as codas and differences in the distribution of fortis release ... would go unexplained' (p. 11).

This is not necessarily so. Although an emergentist theory of morphological structure is unlikely to posit representations such as those above, it is nonetheless capable of distinguishing between the internal structure of verbs that exhibit pseudo-inflection and those that exhibit regular inflection. Consider in this regard the theory of emergentist morphology put forward by Hay (2003), who proposes (based on experimental evidence) that the internal structure of affixed words is gradient rather than categorical, reflecting their relative frequency compared to that of their stem. The words *inadequate* and *inaudible* are a case in point, as I noted in my opening essay. Because *adequate* is more frequent than the affixed form *inadequate*, its presence in the derived word is relatively salient, leading to a high native speaker rating for structural complexity. In contrast, because *audible* is less frequent (and therefore less salient) than *inaudible*, the affixed form receives a low rating for structural complexity.

By this criterion, relatively strong morphological boundaries should emerge for regulars such as *wrapped* and *helped* since in both types of case, the stem (*wrap* and *help*, respectively) is likely to be reasonably frequent in the input (in addition to its presence in instructional materials such as glossaries and dictionaries). In contrast, the 'stem' for the pseudo-inflected form (*kep* in the case of *kept*, *slep* in the case of *slept*, etc.) has a frequency of zero. There is therefore no opportunity for a word-internal morphological boundary to emerge in these cases. As a result, such words end up being monomorphemic.

This in turn opens the door to an alternative explanation for the distribution of fortis release in the speech of the Mandarin learners. In particular, it is possible that this phonetic strategy is used to indicate the presence of the (strong) boundary between a stem and an inflectional affix.[3] Interestingly, a parallel tendency—this time involving native English speakers—has been documented by Hay (2003:123ff) for a different type of morphological boundary. She reports that the stem-final /t/ is pronounced more strongly in *daftly* and *softly* (whose stem is more frequent than the affixed word, creating a strong internal morpheme boundary) than in *swiftly* (which is more frequent than its stem, resulting in a weaker internal boundary).

My suggestion about the L2 facts also permits a prediction that is yet to be tested: fortis release should not be observed on the [p] in monomorphemic words such as *apt* and *rapt*. Rather, these words should behave like the pseudo-inflected *kept* and *slept*, which are also monomorphemic.[4]

Whatever the fate of this prediction—and, indeed, of the entire proposal that I have just sketched—it is clear that an emergentist analysis can and does posit a representational difference between the two types of inflected forms (regular and irregular) while at the same time placing the stem-final consonant in the coda in both types of words, consistent with the phonetic facts. This in turn opens the door to various possible explanations for differences in the L2 pronunciation of the consonant cluster in *kept* and *wrapped* without the need for reference to onset-nucleus sharing, null nuclei, or an innate Empty Category Principle.

3. Concluding Remarks

A theory of language acquisition that incorporates some version of the 'innateness hypothesis' is not the same thing as a theory of language acquisition that incorporates a version of UG. And an emergentist theory of language acquisition is not the same thing as a theory of frequency effects.

The purpose of emergentism is not to refute nativism; it is to devise a better version of the innateness hypothesis. (It's no accident that the title of Elman et al.'s pioneering work is *Rethinking Innateness*, not *Rejecting Innateness*.) That enterprise encompasses many different possibilities, and it is perhaps premature to take a hard position on anything at this time. Nonetheless, it is possible to discern certain promising lines of inquiry, including the idea that a processor that is sensitive to computational efficiency as well as to frequency can shed light on many of the classic mysteries that originally motivated interest in Universal Grammar. Further pursuit of this idea requires careful and detailed attention to the traditional problems

[3]The question of why Chinese speakers employ this particular phonetic strategy is yet to be addressed. However, it is perhaps relevant to note that Mandarin does have aspirated consonants, and that it permits voiceless vowels in unstressed, toneless syllables under certain conditions (Chao 1967: 21-27).

[4]It would also be interesting to know whether the word-final [t] of *apt* and *rapt* manifests fortis release. Goad & White observe that there is a higher rate of fortis release of the [t] in the case of pseudo-inflection than regular inflection (i.e., *kept* vs. *wrapped*).

of language acquisition, including the very important phenomena considered in the intriguing UG contributions to this issue.

References

Bailes, Gregg. 2000. *Wanna* contraction in a British English environment. M.A. thesis. Department of Linguistics and English Language, University of Durham.

Baltin, Mark. 2006. The non-unity of VP-preposing. *Language* 82, 734-66.

Bowerman, Melissa. 1987. Commentary: Mechanisms of language acquisition. In B. MacWhinney (ed.), Mechanisms of language acquisition, 443-66.Mahwah, NJ: Erlbaum.

Bybee, Joan. 2002. Sequentiality as the basis of constituent structure. In T. Givón & B. Malle (eds.), *The evolution of language out of pre-language,* pp. 109–134. Philadelphia: John Benjamins.

Bybee, Joan. 2003. Mechanisms of change in grammaticization: The role of frequency. In B. Joseph & R. Janda (eds.), *The handbook of historical linguistics*, 602-23. Oxford: Blackwell.

Bybee, Joan & Joanne Scheibman. 1999. The effect of usage on degree of constituency: The reduction of *don't* in American English. *Linguistics* 37, 575–596.

Chao, Yuen Ren. 1967. *Mandarin primer* (5th ed.). Cambridge, MA: Harvard University Press.

Chomsky, Noam. 1975. *Reflections on language.* New York: Pantheon.

Chomsky, Noam. 2005. Three factors in language design. *Linguistic Inquiry* 36, 1-22.

Crain, Stephen, Takuya Goro & Rosalind Thornton. 2006. Language acquisition is language change. *Journal of Psycholinguistic Research* 35, 31-49.

Elman, Jeffrey, Elizabeth Bates, Mark Johnson, Annette Karmiloff-Smith, Domenico Parisi, & Kim Plunkett. 1996. *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

Fitch, W. Tecumseh, Marc Hauser, & Noam Chomsky. 2005. The evolution of the language faculty: Clarifications and implications. *Cognition* 97, 179-210.

Goad, Heather & Lydia White. 2006. Ultimate attainment in interlanguage grammars: A prosodic approach. *Second Language Research* 22, 243-68.

Gregg, Kevin R. 2003. The state of emergentism in second language acquisition. *Second Language Research* 19, 95-128.

Hauser, M., E. Newport, & R. Aslin. 2001. Segmentation of the speech stream in a nonhuman primate: Statistical learning in cotton-top tamarins. *Cognition* 78, 53-64.

Hay, Jennifer. 2003. *Causes and Consequences of Word Structure.* New York: Routledge.

Hespos, Susan & Elizabeth Spelke. 2004. Conceptual precursors to language. *Nature* 430, 453-56.

Jusczyk, Peter. 1997. *The discovery of spoken language*. Cambridge, MA: MIT Press.

Krug, Manfred. 1998. String frequency: A cognitive motivating factor in coalescence, language processing and linguistic change. *Journal of English Linguistics* 26, 286–320.

Kuhl, Patricia & J. Miller. 1975. Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. *Science* 190, 69-72.

MacWhinney, Brian. 2002. Language emergence. In P. Burmeister, T. Piske, & A. Rohde (eds.) *An integrated view of language development: Papers in honor of Henning Wode*, pp. 17-42. Trier: Wissenshaftliche Verlag.

Matushanksy, Ora. 2006. Head movement in linguistic theory. *Linguistic Inquiry* 37, 69-109.

Morin, Yves-Charles. 1985. On the two French subjectless verbs *voici* and *voilà*. *Language* 61, 777-820.

O'Grady, William. To appear. Language without grammar. In N. Ellis & P. Robinson (eds.), *Handbook of cognitive linguistics and second language acquisition*. Mahwah, NJ: Erlbaum.

Pinker, Steven & Ray Jackendoff. 2005. The faculty of language: What's special about it? *Cognition* 97, 201-36.

Raums, Franck, Marc Hauser, Cory Miller, Dylan Morris, & Jacques Mehler. 2000. Language discrimination by human newborns and by cotton-top tamarin monkeys. Science, Vol. 288, no. 5464, 349-51.

Rohde, Douglas. 1999. A connectionist model of sentence comprehension and production. Thesis proposal, Carnegie Mellon University.

Slobin, Dan. 1985. Cross-linguistic evidence for the language-making capacity. In D. Slobin (ed.), The cross-linguistic study of language acquisition. Vol. 2: Theoretical issues, 1157-1256.

Tomasello, Michael. 2003. *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Wolfram, W., & Christian, D. 1976. *Appalachian speech.* Arlington, VA: Center for Applied Linguistics.