Assessing Child and Adult Grammar

Julie Anne Legate & Charles Yang
University of Pennsylvania
To appear in Berwick & Piatelli-Palmarini (Eds.)
Rich Languages from Poor Inputs: In Honor of Carol Chomsky

February 8, 2011

1 Introduction

Idealization is the prerequisite for theoretical progress, yet it requires constant revision to keep in touch with reality. The assumption of the child as an instantaneous learner has helped sharpen the focus on the properties of Universal Grammar (UG), though it inevitably deprives us of insights into the process of language acquisition. As Carol Chomsky's pathbreaking research shows, we stand to gain much from the transient stages in child language. Not all aspects of child language are acquired instantaneously or uniformly: acknowledging this in no way denies the critical contribution from UG and can only lead to a more complete understanding of child and adult language. To do so requires accurate measures of children's developmental trajectories, realistic estimates of the primary linguistic data, concrete formulations of linguistic theory, and precise mechanisms of language acquisition. It is in this spirit that we tackle the acquisition of the English metrical stress system in the present paper.

Why stress? First, the stress system of English has played a significant role in the development of phonological theories (Chomsky & Halle 1968, Liberman & Prince 1977, Hayes 1982, 1995, Halle & Vergnaud 1987, Halle 1998) yet considerable disagreement remains. The developmental patterns of stress acquisition may shed some light on the organization of the adult grammar as Carol Chomsky's work demonstrated. Second, there is now a reasonable body of developmental data on stress acquisition, both longitudinal and cross sectional, that the main (early) stages in children's metrical system can be identified–although as we shall see, more studies are still required before the phonological theory of stress can be fully connected with child language acquisition. Third, and quite generally, linguistic theories frequently have to make decisions on what constitutes the core system of the grammar–see, e.g., *basic* word orders, *default* rules, *un-*

marked forms–and what can be relegated to the lexicalized margins. The complex metrical system of English is riddled with exceptions, thanks in part to the extensive borrowing in the history of the language. As far as we can see, theoretical devices that express these idiosyncracies–see, e.g., diacritics, exception marking, or "lexical listing"–are frequently asserted without principled basis. Of course, these are decisions the child learner needs to make as well for the primary linguistic data does not arrived pre-labeled as core or peripheral; the child's the navigation toward the adult grammar might shed light on the choices of linguistic theorizing as well.

Our approach is to deploy a model of learning on the type of data that a young English learner might encounter. The learning model is designed to detect structural productivity, or lack thereof, in the face of exceptions–exactly the type of situation that a metrical stress learner of English faces. We evaluate the validity of generalizations in the metrical system that the learner might arrive at, and we aim to relate these to the developmental stages in child grammar and the theoretical treatments of stress in adult grammar.

2 Learning Productivity

How many exceptions can a productive rule tolerate? Our approach is a throwback to the notion of an *evaluation measure*, which dates back to the foundations of generative grammar (Chomsky 1955, Chomsky & Halle 1968, in particular p172). It provides an evaluation metric, and hence a decision procedure, that the learner can deploy to determine whether a linguistic generalization is productive and thus can be extended to new items that meet its structural description.

Generative grammar traditionally holds that the computation of linguistic forms is governed by the Elsewhere Condition (Anderson 1969, Kiparsky 1973, Halle and Marantz 1993, Halle 1997), which requires the application of the most specific rule/form when multiple candidates are possible. This provides a way for representing exceptions and together with rule-following items. Algorithmically, the Elsewhere Condition may be implemented as a serial search procedure:[1]

(1) IF $w = w_1$ THEN ...
IF $w = w_2$ THEN ...
...
IF $w = w_m$ THEN ...
R

[1]This is also a return to the traditional, following the serial search model for lexical access (Forster 1976, 1992), which in turn can be traced at least back to memory scanning models of Sternberg (1969). The advantage of this model lies in its ready availability for analytic methods, and its straightforward account of frequency effects in lexical processing (Murray & Forster 2004).

The exception clauses (1—m) in (1) can be ordered with respect to their frequencies of occurrence by online algorithms (e.g., Rivest 1976). Thus the time required to access a specific entry among the exceptions will be correlated with its position on the list: more frequent entries will be placed higher on the list and will thus be accessed faster, a result that has been consistently confirmed in the reaction time studies of irregular processing (e.g., Marslen-Wilson & Tyler 1997). The key feature of the model in (1) is that productivity comes at a price: a productive rule may induce considerable time complexity in on-line processing. Specifically, the computation of a rule-following item will have to "wait" until all exception clauses, or $1 - m$ in (1), are evaluated and rejected. Thus, if a rule has too many exceptions, the overall complexity of morphological computation may be higher under (1) than simply listing everything. This would arise if there are too many rule-following items made to wait when their frequencies of occurrence would have placed them higher—and accessed faster—on a complete listing model of processing.

An immediate prediction of the model in (1) is as follows. Take two words, w_e and w_r, the former being an exception to a productive rule R, whereas w_r is a rule following item. Psycholinguistic evidence for the serial search of exceptions before rules dates back at least to Cutler & Swinney (1979), where it is shown that idiomatic expressions such as "kick the bucket" are processed faster than the compositionally formed "lift the bucket" where word and word-transition probabilities are matched. Additional evidence comes from the study of morphological processing. Thus, the model in (1) entails that w_e will be computed faster than w_r *if* the following conditions are met:

(2) a. the lemma/stem frequencies of w_e and w_r are matched, *and*
    b. the frequencies of the rules that w_e and w_r make use of, i.e., the sum of the token frequencies of all words that follow these rules, are also matched.[2]

The familiar case of English past tense, unfortunately, is not applicable here. While the irregular verbs are highly frequent, none of the irregular processes comes close to the total frequency of the productive "add -d" rule, which collects relatively lower frequency regular verbs but in very high volume (Grabowski & Mindt 1995). The most appropriate tests can be found in two pockets of German morphology. The first test comes from the German noun plural system. The default rule is to add an -s suffix, but there are four other classes with varying degrees of productivity (Marcus et al. 1995). Of interest is the decidedly non-productive class that adds the -er suffix, which is closely matched with -s class in rule frequency (e.g., Sonnenstuhl & Huth 2002). Lexical decision studies show that when -r and -s suffixing nouns are matched in stem frequency,

[2] A frequently used rule will be processed faster than a less frequently used one; this effect can be observed in both morphological processing (Sereno & Jongman 1997) and morphological learning (Yang 2002).

the -s words show considerably slower reaction time (Penke & Krause 2002, Sonnenstuhl & Huth 2002). Another case concerns the formation of past participles in German, where the default rule is to use the -t suffix and there is an unpredictable set of irregulars that add -n. Despite the low type frequency of -n verbs, "add -t" and "add -n" classes are comparable in rule frequency. In an online production study, Clahsen et al. (2004) find that when stem frequency is controlled for, the regular "add -t" class is slower than the irregular classes, at least for words in the higher frequency region which normally constitute the basis for productivity calculation during language acquisition.

Processing considerations based on time complexity provides us with a cost-benefit calculus on productivity. Suppose that there exists a rule R that can in principle apply to a set of N lexical items. Of these, m items are exceptions and they are represented in the form of the Elsewhere Condition (1). Let $T(N, m)$ be the expected time of rule access if R is productive: in other words, $(N - m)$ items will need to wait until the m exceptions have been searched and rejected. By contrast, if R is not productive, then all N items must be listed as exceptions, again ranked by their frequencies; let $T(N, N)$ be the expected time of rule access for a list thus organized. We conjecture:

(3) **Tolerance Principle**: R is productive if $T(N, m) < T(N, N)$; otherwise R is unproductive.

The reader is referred to Yang (2005) for the mathematical details of the model. In essence, for the $(N - m)$ well behaving items, the rule search time is the constant m, the number of exceptions that must be ruled out. For an item in the set of m exceptions, the rule search time is its rank/position on the list. Thus the expected time can be expressed as the rule search time weighted by frequencies. Assuming that the items in the exception clauses in (1) follow the Zipfian distribution, it is possible to show that:

(4) **Theorem**: R is productive if and only if

$$m < \frac{N}{\ln N}$$

That is, the number of exceptions would need to be fairly small compared to the number of rule following items to warrant productivity.

The Tolerance Principle can be straightforwardly applied to identify both productive and unproductive processes in languages. The case of English past tense is obvious: supposing that there are 120 irregular verbs, one needs a total of 800 ($800/\ln 800 \approx 120$) verbs altogether, or 680 regulars, to sustain the productivity of the -d suffix, which is of course easily met. Take another well known case in the psycholinguistic study of morphology: the plural formation of nouns in German briefly discussed above. The failure of the Tolerance Principle would be total if puralization in German operates as claimed in some quarters (e.g., Marcus et al. 1995) with only one productive rule ("-s"), which accounts for only a tiny fraction of nouns (about 5%; Sonnenstuhl & Huth 2002): the -s rule would have 5% coverage and 95% exceptions. Thus there must be productive processes within the so-called irregulars. One quickly discovers that the feminine nouns in German tend to take the -n suffix though all grammatical descriptions are quick to point out the existence of a considerable number of feminine nouns that take other suffixes. The Tolerance Principle can be used to evaluate these generalizations. For monomorphemic[3] feminine nouns that have appeared at least once per million in the Mannheim corpus, 709 take the -n suffix while 61 do not—which is well below the tolerance threshold of $770/\ln(770) \approx 116$. Thus, the -n suffix is predicted to be productive for feminine nouns. Two converging lines of evidence support this prediction. First, German children overuse the -n suffix as frequently as the -s suffix (Szagun 2001): the two thus must both be productive, which is the prerequisite for over-regularization. Second, lexical decision tasks show no whole-word frequency effect among the -n suffixed nouns—a hallmark for productive word formation processes (Penke & Krauss 2000). The claim of a productive -n rule has been made by many specialists on German morphology (Wiese 1996, Wunderlich 1999), often in reaction to the dual route position of Marcus et al. (1995). The novelty of the present approach lies in its ability of reaching similar conclusions on purely numerical basis.

Under the Tolerance Principle, mere majority of a form does not entail productivity; only a filibuster-proof super majority will do, as the sublinear function $1/\ln N$ translates into a small number of exceptions.[4] Another case in English past tense illustrates the opposite side of productivity: paradigmatic gaps. It is well known (e.g. Pinker 1999; see also Gorman 2012) that the irregular stem *forgo* has no generally accepted past tense form (**forwent*, **forgoed*) while *stride* has no generally accepted past participle form (**strided*, **striden*). Following the original discussion of such matters (Halle 1973, in particular footnote 1), these ineffable forms can only arise in the unproductive regions of word formation, for otherwise a productive rule would automatically apply (as in the case of the *wug* test). Suppose the learner has encountered a verb for which the past tense or past participle form is irregular, i.e., not the regular -d form. He now knows *undergo* and *stride* must be irregular but has not encountered the past tense of the former or the past participle

[3]This is the most conservative estimate. If one includes compound nouns, the number of -n suffixed feminine nouns greatly increases. We thank Kyle Gorman for verifying these counts.

[4]Clearly, none of the English irregular rules can be productive since each would have thousands of exceptions (i.e., regular verbs); this is clearly reflected in the virtually total absence of over-irregularization errors (e.g., *bring-brang*) in child English and other languages (Xu & Pinker 1995, Clahsen 1999).

of the latter. He may also notice the pattern among the irregular verbs that a majority of them have identical forms for the preterite and participle (e.g., *hold-held-held*, *think-thought-thought*). Indeed, in CELEX English lexicon, 102 out of the 161 irregular verbs follow this pattern of syncretism, but the 59 exceptions (e.g., *break-broke-broken*, *sing-sang-sung*) prove fatal. For a set with $N = 161$ items, a valid generalization can tolerate no more than $(161/\ln 161 \approx 32)$ exceptions, which is considerably fewer than the actual number of exceptions. Thus, even though the preterite-participle identity pattern holds for almost twice more items than exceptions, it fails to reach the productivity threshold. We correctly predict that the learner will be at a loss when he needs to "undergo" in the past or "stride" in the past participle.

The application of the Tolerance Principle critically depends on the composition of the vocabulary—or syntactic constructions, see Yang (2010)—that resides in the individual learner. The productivity of a certain process may even change, along with its scope of application and exceptions—the two quantities N and m may fluctuate as the learner processes more primary linguistic data. We return to these issues in the acquisition of the metrical stress system of English. To begin, let us consider the adult language system that serves as the target of acquisition.

3 A Sketch of English Stress

The English stress system is complex enough to have engendered a number of competing theoretical analyses, though several points of generalization are common to most. Roughly speaking, main stress in the nominal domain falls on a heavy penult, and otherwise on the antepenult. In verbs, main stress falls one syllable closer to the word boundary: on a heavy final, and otherwise on the penult. At a closer level of detail, considerable disagreements remain; here we consider only three alternatives.[5]

Halle & Vergnaud's (1987) system makes use of the bracketed metrical grid notation. For nouns, the system is designed to catpure the generalization that main stress falls: (i) on the final syllable if it contains a long vowel, (ii) on the penult if its rime is branching, and otherwise (iii) on the antepenult. To this end, they propose the following primary components. Final syllables are extrametrical if they contain a short vowel, which prevents final main stress when the final vowel is short. The quantity sensitivity of the system is encoded through branching rimes projecting an asterisk on Line 1. These asterisks are respected while working from right to left, Line 0 is organized into bounded (unary or binary) feet, whose leftmost element projects an asterisk on Line 1. To ensure that the rightmost of these elements bears main stress, line 1 is organized

[5]The work of Burzio (1994) is considerably different from those reviewed here. Although it contains interesting insights and an alternative perspective, we do not find it sufficiently mechanistic to allow for concrete evaluations.

into a single unbounded foot, whose rightmost element projects an asterisk on Line 2. Further complications aside, this Line 2 asterisk locates the main stress.

				x	Line 2
		(x		x)	Line 1
(5)	a.	(x	x)	(x)	Line 0
		kan	ga	roo	
			x		Line 2
		(x	x)		Line 1
	b.	(x)	(x)	<x>	Line 0
		hor	i	zon	
		x			Line 2
		(x)			Line 1
	c.	(x	x)	<x>	Line 0
		Ca	na	da	

In verbs, stress generally falls on one of the final two syllables. Since antepenultimate stress is not at issue, extrametricality is not posited for verbs. Furthermore, the determination of quantity sensitivity is different for verbs, in that a word final consonant is ignored. Otherwise, the system described above for nouns applies, yielding results wherein stress falls on a final superheavy syllable, and otherwise on the penult:

			x		Line 2
		(x	x)		Line 1
(6)	a.	(x)	(x	x)	Line 0
		con	si	der	
				x	Line 2
		(x	x	x)	Line 1
	b.	(x)	(x)	(x)	Line 0
		in	tro	duce	

Hayes (1995; also 1982) represents another significant approach in the theory of metrical stress, notable by the diversity of stress systems surveyed there. Its treatment of English, however, is most detailed when considering diagnostics for stress and stress shifts above the level of the word. It lists English as a language with "Latin-like" stress, a system that he analyses as follows. The final syllable is extrametrical, and feet are trochaic, where the trochee may consist of a heavy syllable followed by a light syllable, or a heavy syllable alone. Heavy syllables consist of a short vowel with one or more coda consonants, or a long vowel (number of coda consonants immaterial). Thus, when the penult is light, a bisyllablic trochee will be formed, whereas when the penult is heavy a monosyllabic trochee is formed. Again, the rightmost strong element receives main stress at the word level. In addition, as in Hayes (1982), the 1995 approach claims that material smaller than the syllable may be extrametrical, notably the segment. Thus, a word-final consonant in English is also extrametrical, which has an effect on the main stress of verbs but not nouns, since for nouns the entire final syllable is extrametrical. This results in representations of English that are identical to Halle & Vergnaud in the relevant respects.[6] Hayes (1982) also considers the issue of word-final stress in English nouns. He asserts that nouns with a long vowel in the final syllable "always" receive (primary or secondary stress), e.g. *monsoon, misanthrope*. Nouns with a short vowel in the final syllable may exceptionally be stressed, e.g. *gymnast* though Hayes does not provide an example of final primary stress on a short vowel. Thus, again, the extrametricality rule for English nouns cannot apply to final syllables containing a long vowel, and nouns with final stress on a short vowel will be lexical exceptions. To the best of our knowledge, here as well as elsewhere in the metrical stress literature, no principled account of what counts as lexical exceptions is on the offer.

The approach in Halle (1998) departs from previous approaches in a number of respects. An innovation that he presents as central, though we do not dwell upon, is the assumption from Idsardi (1992) that foot construction rules insert left/right foot brackets, rather than building a foot. This has the effect that the left and right brackets of a foot may be inserted through two distinct rules. It also requires an interpretation for mismatched parentheses: these are treated as closed at their maximal size. Thus, (7a) and (7b) are treated as (7c), and (7d) and (7e) are treated as (7f).

(7)
a. (x x x x x
b. x x x x x)
c. (x x x x x)
d. (x x x (x x x)
e. (x x x) x x x)
f. (x x x) (x x x)

Rather than positing extrametricality per se, Halle proposes two edge marking rules for nouns. The first inserts a right bracket between a word-final and a penultimate asterisk, where the word-

[6] Hayes' work is also notable for the claim that "English stress is both listed in the lexicon *and* derived by rule" (1982:237). It is not clear to us why the language learner would duplicate their effort in this way, nor how one may evaluate this claim.

final asterisk is the projection of a short vowel. This ensures that a final syllable containing a short vowel will not be part of the foot containing main stress (cf. Hale & Vergnaud's extrametricality of a final syllable containing a short vowel). When the word-final asterisk is the projection of a long vowel, a left bracket is inserted instead; this parses the final syllable into a unary foot, which will later allow it to bear secondary stress. The main stress rule itself consists of two subrules that refer to the brackets inserted by these edge-marking rules. The first subrule inserts a right bracket two syllables left of: (i) either of these brackets, or (ii) the end of the word. It applies only when the first syllable to the left of the bracket or word boundary is light. Given that line 0 feet are left-headed, for nouns with a short vowel in the final syllable, and a light penult, this results in antepenultimate main stress:

	x			Line 1
(8)	(x	x	) x	Line 0
	Ca	na	da	

The second subrule for main stress assignment inserts a right bracket one syllable left of: (i) either of the brackets inserted by the edge-marking rules for nouns, or (ii) the end of the word. This results in penultimate main stress for nouns with a short vowel in the final syllable and a heavy penult:

	x	x		Line 1
(9)	x	(x	) x	Line 0
	hor	i	zon	

For nouns with a long vowel in the final syllable, two feet project an asterisk to line 1. To adjudicate between them, Halle posits a Rhythm Rule, which inserts a right bracket before the initial asterisk on Line 1; noteworthy is that the unbounded foot thus created is left headed. Hence, primary stress again falls on a heavy penult and otherwise on the antepenult. As the bearer of a line 1 asterisk, the final syllable receives secondary stress.[7]

			x		Line 2
			(x	x	Line 1
(10)	a.	x	(x	(x	Line 0
		sta	lag	mite	

[7] We leave aside Halle's treatment of secondary stresses that precede the main stress; see Hale (1998, p554) and Halle & Kenstowicz (1991).

```
      x                 Line 2
      (x           x    Line 1
b.    (x    x     (x    Line 0
      ma    la   chite
```

Turning to verbs, again, since antepenultimate stress is not at issue, the edge-marking rules are inapplicable. The two subrules composing the main stress rule apply identically, though, providing for main stress on a heavy final and otherwise on the penult. Although Halle does not discuss what constitutes a heavy syllable, it is clear from his examples that, like Halle & Vergnaud and Hayes, a long vowel or short vowel and coda consonant are sufficient to create a heavy syllable, with the provision that a word-final consonant is discounted (at least for verbs). For example, the penultimate stress on *develop* is achieved by considering the final light.

```
                  x           Line 1
(11)   a.   x    (x     x     Line 0
            de    ve   lop

                        x     Line 1
       b.   x     x    (x     Line 0
            in   tro   duce
```

An additional feature of Halle's system is the proposal that certain English suffixes are wholy or partially *unstressable*, by which he intends, when they are word final, they are ignored by the above rules of stress assignment. These include *-y*, the final syllable of *-ory/ary*, *-ive, -ton, -shire, -er,*. We do not directly address this issue in our consideration of learning below but will point to

In sum, major differences between the models arise largely in the treatment of nouns with long vowels in the final syllable. In the non-exceptional case, Halle & Vergnaud predict final primary stress, Hayes is indeterminate between final primary or secondary stress, and Halle predicts final secondary stress, except in the case of a final long unstressable syllable, which will not bear stress.

4 The Learning Model

We assume that the child learner has acquired a sufficient amount of phonological knowledge of her specific language to carry out the computation and acquisition of metrical stress. Specifically, we assume

(12) a. That the child has acquired the segmental inventory of the native language, which is typically fairly complete before her first birthday, even though the mechanisms by

which such learning takes place are currently unknown (Werker & Tees 1983, Kuhl et al. 1992; see Yang 2006 for review).

b. That the child has acquired the basic phonotactic constraints of the language (Halle 1979) and thus capable of building syllables from segments which are subsequently used to construct the metrical system.[8] For instance, Dutch and English learning infants at 9 month prefer consonant clusters native to their languages despite the segmental similarities between these two languages (Jusczyk et al. 1993).

c. That the child is capable of extracting words from continuous speech, perhaps as early as seven and half months (Jusczyk & Aslin 1995). While the role of statistical learning in word segmentation (Saffran, Aslin & Newport 1996) is not useful as previously thought, universal constraints on lexical stress (Halle & Vergnaud 1987, Yang 2004) and the bootstrapping use of previously segmented words (Jusczyk & Hohne 1997, Bortfeld et al. 2005) appear to be sufficient for the task of segmentation, at least for English (Yang 2004).

d. That the child can readily detect prominence of stress. Indeed, very young infants appear to have identified the statistically dominant stress pattern of the language, as 7.5 month old English learning infants perform better at recognizing trochaic than iambic words (Jusczyk, Cutler & Redanz 1993, Jusczyk, Houston & Newsome 1999): at the minimum, the child is able to locate primary stress on the metrical structure of words, and acquisition of the metrical system probably starts well before the onset of speech. We return to the issue of trochaic preference in early child language, as it appears to be a transient stage toward the target grammar.

These assumptions are warranted by the current understanding of prosodic development in children and appear indispensable for any formal treatment of stress acquisition.

We share the insights emerging from metrical theories (HV, Hayes, Idsardi) that stress acquisition can be viewed as an instance of parameter setting as the learner makes a set of choices made available by UG. However, we part ways with previous efforts on metrical stress acquisition in the following ways. Unlike Tesar & Smolensky (2000) and much of the acquisition research in Optimality Theory, we do not assume that the learner has access to target-like representation of the metrical structure, which would largely trivialize the learning process. Indeed, similar complaints may be lodged against all learning models that provide the learner with both the underlying and surface representations of linguistic data: recovering the underlying structure from

[8]See Gorman (forthcoming) for a modern assessment of the extent to which phonotactics can be regarded as a consequence of phonological knowledge as the traditional position holds (Halle 1962), rather than an independent component of grammar.

the surface structure is the task of the grammar, the very target of learning.[9] In addition, the criticisms lodged at the cue-based approach below, in particular the issue of productivity and exceptions, apply equally to OT and corresponding learning models: the data does not go away under constraints.

In what is known as the cue-based learning approach (Dresher & Kaye 1990, Dresher 1999),[10] the metrical parameters are set in an ordered sequence, each of which is crucially conditioned upon the choices of prior decisions. For instance, while syllables containing a long vowel (VV) may universally be regarded as heavy and syllables with a short vowel without coda (V) light, the weight of those with short vowel and coda consonants (VC) is a choice of the *rime parameter* for the specific language. However, the rime parameter is only "active" for metrical systems, as in English, that are *quantity sensitive*, where the stress placement makes crucial reference to syllable weight. Languages such as Maranungku are, by contrast, *quantity insensitive*: the primary stress falls on the initial syllable, and secondary stresses on every odd syllable thereafter regardless their weights. Thus, the quantity sensitivity parameter must be set prior to the rime parameter, which likewise must precede the setting of the stress placement parameters.

A major motivation for learning as a sequence of decisions is to uphold the idealization of the child as a deterministic learner. For instance, suppose the child has not yet determined the quantity sensitivity of the language: if he proceeds to the stress placement parameters in a quantity sensitive language such as English, he might as well need to retreat from these parameters. But this idealization of deterministic learner is both empirically problematic and formally unnecessary. As we shall see, there *is* an initial stage of stress acquisition of Dutch (Fikkert 1994), a quantity sensitive language, that can appropriately characterized as quantity insensitive (cf. Kehoe & Stoel-Gammon 1997), and the child does seem able to backtrack from this incorrect hypothesis before heading toward the target. Moreover, with the advent of UG-based probabilistic learning such as the variational model (Yang 2002, Straus 2008), the formal learnability motivations for cues are no longer necessary. Consider two parameters A and B, where the correct value of B can only be determined after the value of A. Under cue-based learning, these two parameters will be innately associated with specific types of data *a* and *b*. The learner will first look for *a* to set parameter A before turning its attention to *b* and B. Under the variational model, both A and B are associated with two probabilities that denote the likelihood of their target values, and the learner probabilistically selects the values of A and B to analyze the input data. Data of the type

[9]Conceivably, a joint inference approach could be used to infer both the underlying structures and the grammar mapping them to surface structures, which the learner can directly observe from the input. However, these techniques, which have been used in natural language processing, rely on supervised training methods, and we are not aware of any successful application in models of language acquisition.

[10]See Baker (2001) for a similar approach in syntax.

a will gradually push A toward to the target value; during this process, there is no guarantee that the parameter B will also be pushed toward the target even if the learner encounters b. However, as A moves closer to the target, i.e., the probability of the learner using the target value of A gets higher, b will become increasingly more effective in pushing B toward the target value as well. It is easy to see that eventually both A and B will converge to the target values. Thus one does not need to suppose any ordering of these parameters, nor any innate association between $A - a$ and $B - b$: the *presence* of a and b alone is sufficient to guarantee learnability.[11]

Abandoning the idealization of a deterministic learner also has the benefit of bringing the abstract model closer to the reality of language learning. The formulation of the cue-based learner requires the learning data to be "sanitized" (Dresher & Kaye 1990) as to steer the child clear from noise, lexical exceptions and the like. However, the learning data should be realistic, and thus must contain both positive instances exemplifying the target grammar as well as lexically and/or structurally conditioned exceptions. On the other, one cannot uncritically assume the ready availability of especially informative items in the input; the welfare of the child's metrical stress should not be left to chance—needing to hear *Manitoba* or *Winnipesaukee*, for instance.

More important, and more general to the theory of language and language learning, is the issue of balancing generalizations with exceptions. In more recent treatments of cue-based learning (Dresher 1999), it was recognized that the learner's choice may be influenced by the composition of the linguistic data. For instance, if the child were to suppose that English has a quantity insensitive stress acquisition, then words with n syllables must be stressed consistently. Dresher points to the presence of a few counterexamples to this conjecture (e.g., *América* but *Minnesóta*) as cues for the child abandoning quantity insensitivity. However, this approach would disqualify all generalizations about English stress as every theory must deal with the exceptions. The learner's dilemma reduces to that of productivity: quantity insensitivity may be upheld if the patterns such as *América* and *Minnesóta* are not sufficiently abundant and can be listed as lexical exceptions.

Thus, the productivity model outlined in section 2 will play a critical role in our approach to metrical stress. While the reaction time studies provide direct evidence for the conception of rules and exceptions (1) in morphology and syntax, we are not aware of similar cases in phonology. The preliminary success of the model reviewed in section 2, and reported in comprehensive details in Yang (in preparation), provides us with sufficient motivation for its applicability in the present case. We outline our approach below.

[11]In the most comprehensive study of a syntactic parameter space (Fodor & Sakas, submitted), it seems that only a small number of parameters have the sort of conditional dependency conceived in the cue-based approach; the majority of parameters can be set completely independently. This strengthens the learnability result of the variational learning model.

Universal Grammar provides a core set of parametric options that delimit a range of possible metrical structures (syllable, weight feet) and possible computational operations (e.g., projection, foot building, edge marking) that manipulate these structures. Frequently the stress rules are subject to highly language specific structural conditions beyond the metrical system; see, as reviewed in section 3, the stress patterns for nouns as opposed to verbs in English (see also Roca 2005 for Spanish), and a variety of affixes with stress shifting properties. It is inconceivable that the *totality* of these options is available to the learner. Rather, we envision the learner experimenting and evaluating the core metrical hypotheses in an incremental fashion as he processes linguistic data, and the learner chooses the grammar most highly valued with respect to the present data:

(13)
a. If a grammar fails to reach productivity as prescribed by the Tolerance Principle (4), it is rejected.
b. If there are multiple grammars meeting the Tolerance threshold, the learner selects the one with fewest exceptions (i.e., most productive).
c. If no grammar is productive, then the stress patterns of words are memorized as a lexicalized list.[12]

Each grammar G_i, then, can be associated with a tuple (N_i, m_i), the number of words (N_i) it could apply to, and the number of words that contradict it (m_i). Thus, the learner traverses through a sequence of grammars as learning proceeds, presumably reaching the target G_T in the end:[13]

(14) $G_1 \rightarrow G_2 \rightarrow G_3 \rightarrow ... \rightarrow G_T$

Under this view, G_{i+1} is more highly valued than G_i resulting from additional linguistic evidence unavailable at the stage of G_i. In particular, the additional data may have the effect rendering G_i unproductive thereby forcing the learner to adopt a different grammar G_{i+1}.[14]

In general, it is possible that a grammar's productivity changes as learning proceeds; after all, the numerical basis of productivity (N_i and m_i) changes as the child learns more words. The

[12]This is not to say that the learner directly memorizes the stress pattern of words. If the acquisition of morphophonology is of any relevance, it seems that the learner would uses rules to generate the stress patterns of words—it's just that these rules are not productive. See Yang (2002) for such a treatment of the English irregular verbs, in contrast to the direct memorization approach in the dual-route morphology literature (Pinker 1999).

[13]Strictly speaking, of course, there is no target grammar for the learner converges to. The learner reaches a terminal state, his I-language, based on the linguistic data he receives during language acquisition. Since the data is necessarily a sample of the environment, it is possible that the learner converges to a grammar that is distinct from that of the previous generation of learners, thereby leading to language change. See Yang (2011) for an application of the productivity model to the well known case of noun/verb diatonic stress shift in the history of English.

[14]This process of learning, which we believe is what Chomsky put forward in *Aspects* (1965), is somewhat different in character from the acquisition process in syntactic learning, perhaps reflecting the differences between phonological and syntactic systems (Bromberger & Halle 1989). For additional discussion, see Yang (2010).

well-known U-shape learning in English past tense is a case in point (Marcus et al. 1992). The child initially use irregular verbs correctly[15] (*hold-held*) before succumbing to over-regularization (*hold-holded*), signaling the "-d" rule has reached its threshold of productivity. Prior to this, then, the child would be treating regular verbs (*walk-walked*) as if they were irregular; presumably he has not acquired enough regular verbs for the "-d" rule to counter the irregular verbs, which tend to be more frequent and are more likely to be acquired early. The productivity model can make precise quantitative predictions about the tipping point of rules; the reader is directed to Yang (2005, 2010) for extended discussion.

It is also possible that UG provides certain markedness hierarchies, which lead the learner to entertain some grammars before others. For instance, it is conceivable that quantity insensitive systems are simpler than quantity sensitive ones, and the learner will evaluate the latter only if the former has been rejected by the linguistic data. Alternatively, the learning mechanism may consist of simplicity metric—e.g., the length of the grammar (Chomsky 1955)—that favors certain grammars over others. And all such constraints can be construed as categorical principles or stated in a probabilistic framework of learning.

To operationalize the conception of learning in (14), we will first construct an approximate sample of the child's vocabulary and then evaluate several leading treatments of the English metrical system reviewed in section 3. This exercise serves the dual purpose of testing on the one hand the plausibility of a productivity-driven learning model, and on the other, the descriptive adequacies of theoretical proposals.

5 The Learning Process

To get a realistic assessment of the linguistic input, we took a random selection of about 1 million utterances from child-directed English in the CHILDES database. We approximate the growth of the learner's vocabulary, which serves as the raw material for grammar learning, by extracting words within two frequency ranges to reflect the development of the metrical system.

Since the child directed speech is pooled from multiple children, the resulting sample contains a disproportionately high number of very frequent proper names; although several high frequency proper names (of the child, siblings, or family pets) are almost surely a permanent fixture of each child's life, the totality of the pooled names are not. With the use of an automatic part-of-speech tagger based on Brill (1995),[16] we therefore exclude all proper names from further

[15]To the extent that they mark tense, as past tense learning overlaps with the Optional Infinitive stage in the acquisition of English (Legate & Yang 2007).

[16]Available at http://gposttl.sourceforge.net/.

consideration. In total, 4.5 million words are used for a total of about 26,700 distinct types. We only evaluate the words that have been automatically tagged as nouns and verbs, about 20,000 in all, which constitute the majority of the child's vocabulary for any frequency range. Since nouns and verbs have somewhat different stress patterns, considering them together will pose a realistic test for any model that seeks systematic regularity amidst a heterogeneous mix of patterns.

In some of the studies we describe below, for reasons that will become immediately clear, words are morphologically processed using a computerized database from the English Lexicon Project (Balota et al. 2007) as morphology is also known to play an important role in the computation of stress and it is worthwhile to explore its implications in acquisition. Based on the consistent developmental evidence that the inflectional morphology is acquired relatively early—in some languages very early—we assume that the learner is capable of parsing words into morphological structures and considering their roles in the acquisition of stress.

In all our studies, the computerized pronunciation dictionary CMUDICT version 0.7 is used to obtain the phonemic transcriptions of words, which are then syllabified following the Maximize Onset principle (Kahn 1976) with sonorants and glides in the coda treated as syllabic.[17] We ignore the prosodic effects on lexical stress in the present study. We assume that syllables containing long vowels (diphthongs and the tense vowels /i/ and /u/) are heavy (H), syllables containing short vowels and no coda are light (L); it is the learner's task to determine the proper treatment of syllables with short vowels and at least one coda consonant (C), which may be treated as either H or L depending on the language. For the present paper, we only consider the placement of the main stress. Since the pronunciation dictionary marks primary, secondary, as well as no stress, we mark the former as 1 and collapse the latter two as 0. For instance, the word *animals* will be represented syllabically as `LLC` with the stress contour of `100`.

A thorough assessment of the learning model as encapsulated in (14) would involve an incremental growth of the learner's vocabulary (via Monte Carlo sampling, for instance) and the evaluation of alternative grammars along the way. For simplicity, we only consider two specific points of stress development, one designed to capture the child's stress system under a very small vocabulary and the other when the child has already learned enough words to potentially match the target state.

In the first study on early stress development, we extracted words that only appear more than

[17]Entries that could not be found in these lexical databases are omitted. These are almost exclusively transcription errors or nonsense words in the CHILDES database.

A technical note regarding the utility of electronic databases in the present study. The CMU pronunciation dictionary does not contain part of speech information, making it impossible to distinguish the homographic words with distinct stress patterns (e.g., *recórd* the verb and *récord* the noun.) Words in the CELEX database do contain parts of speech but their phonemic transcription has systematic inaccuracies. We combined the two databases to obtain the correct transcription.

once per 10,000 words, resulting in 420 words most of which, as expected, are relatively simple. The distribution of stress patterns is summarized in Table 1:[18]

contour	counts
1	287
10	107
100	13
01	7
010	3
1000	3

Table 1: Stress patterns for words with frequency $\geq$ 1 in 10,000.

The distribution in Table 1 is clearly consistent with a quantity insensitive trochaic system. A total of 402 words can tolerate $402/\ln 402 = 67$ exceptions where in fact there are 26. Interestingly, children learning English and similar languages go through an initial stage, which terminates at about 2;0, during which the child is limited to a maximum bisyllabic template with the primary stress falling on the first.[19] In the most detailed longitudinal study of stress acquisition, Fikkert (1994) notes that children acquiring Dutch, a language with similar metrical properties as English frequently stress the initial syllable in disyllabic words for which the primary stress falls on the final syllable (e.g., *ballòn→bàllon*, *giràf→gìraf*). Moreover, the few trisyllabic words are invariably reduced to a bisyllabic form, with the primary stress always preserved (e.g., vakàtie→kàntie, òlifant→òfant). Similar patterns have been observed for English learning children (Kehoe & Stoel-Gammon 1997) in a word imitation task.

The preference for a trochaic stress system is not surprising since it is well known that English children's early language has a large number of nouns (Tardif, Shatz & Naigles 1997), most of which are bisyllabic thus heavily favoring the trochee. Of course, the English stress is *not* quantity insensitive, and there are further complications with respect to lexical category and morphological structures. Indeed, if we expand the vocabulary for learning, with more verbal forms coming in, the initial trochaic grammar starts to break down, prompting the learner to develop alternative grammars. To this end, we consider now words that appear at least once per million in our sample of child-directed English, again focusing only on nouns and verbs. There

[18]These extraordinarily long words are *everybody*, *anybody* and *caterpillar*.

[19]Fikkert provides evidence, noted immediately below, for this limitation. Also compatible with our model would be for the child to not be limited to a bisyllabic template, but rather for the child to conjecture a quantitative insensitive grammar with the primary stress on the initial syllable. This grammar is obviously productive, having even fewer exceptions than that discussed in the text.

are 4047 nouns, 2402 nouns, and 5763 lexically and prosodically distinct words altogether.[20]

Now the bisyllabic trochaic grammar drops below the productivity threshold: while still the numerical majority, there are 2388 monosyllabic words and 2145 bisyllabic words with initial stress. A total of 4533 is well below the requisite amount for productivity ($5763/\ln 5763 = 5097$). Even a grammar that is not subject to the two syllable limit and one that always places the primary stress on the initial syllable fails to rise to the occasion. Even though it accounts for an overwhelming majority of words (4960, or 86%), there has been no report of an initial stress strategy in the later development of the metrical system: we take this to be a non-trivial result of the productivity model.

The child, then, must seek alternatives–in the direction of quantity sensitivity, an option in the metrical system. Here the learner has several moves to make. One possibility is to discover regularities within separate lexical classes, e.g., nouns and verbs. Language learning children are well prepared to undertake this task, as the knowledge of lexical categories is acquired extremely accurately (see, e.g., Valian 1986). Another possibility is to consider the interaction between morphology and stress: in English, the inflectional suffixes do not trigger stress shifts in the stems but some of the derivational affixes do (e.g., *-ic* but not *-ment*). This case merits some discussion.

An English learning child is well positioned to take inflectional morphology into consideration in the computation of stress. All inflectional suffixes are learned before 3;6 when measured by Brown's 90% obligatory usage criterion in production, and it is likely that these suffixes are reliably put into use in comprehension even earlier: children as young as 20 month to 2 years old can interpret the inflected verbs of words (Golinkoff et al. 1987) including novel ones (Naigles 1990). Derivational affixes, however, are an altogether different matter. While we do not subscribe to the commonly held view that inflectional and derivational morphologies reflect fundamentally different aspects of grammar (see also Halle 1973), the fact remains that derivational morphology is learned relatively late, perhaps well into the school years (Tyler & Nagy 1989), which may simply be the result of derivational forms are less frequent in the input data and thus providing the learner with fewer instances of data for acquisition. Taken together, we assume that the learner is capable of relating inflectional forms of verbs to their stem forms, but is incapable of parsing derivational forms into decomposable pieces (words such as *growth* and *government* will be treated as morphologically simplex). Furthermore, we assume that the learner has correctly

[20]For words that appear in the input as both nouns and verbs such as *walk* and *record*), they contribute to both the noun and the verb counts; these will be used when the learner evaluates distinct grammars for nouns and verbs. In the case of *walk*, the word only contributes once to the total count of words since the noun and verb form of *walk* are metrically identical. A word like *record*, by contributes, counts twice in the total word counts, since the verb and noun forms of the word are distinct.

learned that inflectional suffixes do not trigger stress shift–a task easily accomplished, again, by the use of the productivity model: there are no exceptions to the lack of stress shift with inflectional morphology. In other words, the child treats all inflectional forms of *walk* (i.e., *walk*, *walks*, *walked* and *walking*) as *walk* for the purposes of stress acquisition. And we return at the end of this section to how the learner may acquire the stress-shifting properties of derivational suffixes.

Following the review in section 3, we compare the placement of primary stress under the Halle & Vergnaud (1987) and Halle (1998) proposals, which are summarized operationally as follows:

(15) The Halle & Vergnaud (1987) system (HV87)

a. Nouns:
   - If the final syllable contains a long vowel (VV), it receives primary stress.
   - Otherwise if the penult is heavy (i.e., VV or VC^+, short vowel with at least one consonant coda), then the penult receives primary stress.
   - Otherwise the antepenult receives primary stress.

b. Verbs:
   - If the final syllable is super heavy (i.e., VV or VCC^+, a short vowel with at least two consonants in the coda), then the final syllable receives primary stress.
   - Otherwise the penult receives primary stress.

(16) The Halle (1998) system (H98):

a. Nouns:
   - If the penult is heavy (i.e., VV or VC^+), then it receives primary stress.
   - Otherwise the antepenult receives primary stress.

b. Verbs: Same as HV87 above (15b).

Table 2 below summarizes the results of evaluating HV87 and H98 under a variety of conditions with respect to inflectional decomposition (**stem**±) and lexical separation (**lex**±). When evaluating grammars without making the lexical distinction ([**lex+**]) between nouns and verbs, we use the noun rules in the HV87 and H98. Since the vocabulary consists of far more nouns than verbs, the failure of the noun rules to reach productivity entails the failure of the verb rules. When evaluating grammars with separate rules for nouns and verbs, we only consider a grammar to be successful if its rules reach productivity for both nouns and verbs. The raw data can be found in the appendix.

lex	**stem**	HV87	H98
-	-	no	no
-	+	no	no
+	-	no	yes[a]
+	+	no	**yes**[b]

Table 2: Evaluation of stress grammars for words with frequency ≥ 1 per million. a. with 515 exceptions. b. with 355 exceptions.

The H98 system under (**lex+**, **stem+**) is declared winner; while H98 under (**lex+**, **stem-**) also manages to reach productivity, it accumulates more exceptions. Unfortunately, there are no direct studies of the interaction between inflectional suffixes and stress–or lack thereof, to be precise–from the transient stages of metrical acquisition, although our results do support the H98 description of the target grammar.

It is interesting to examine the nature of the exceptions under the H98 system, which reveals some interesting patterns considered in Halle's discussion, as well as the traditional literature. Many exceptions in nouns are those with a final syllable containing a long vowel, which ought to receive final stress but do not. Upon inspection, most of these ends in the long vowel /i/, including the final derivational suffix (e.g., the dimunitive *-y/ie* such as *kitty* and *doggie*) as well as morphologically simplex words such as *body* and *army*. Halle notes (see also Liberman & Prince 1977) that these suffixes are unstressable and are therefore ignored by the rules for stress assignment. Although he does not address how the learner might reach such conclusions, the productivity model can be straightforwardly deployed for this task. The morpheme segmentations in the English Lexicon Project lists 530 words with -y suffix: none receives primary stress, or even secondary stress. The productivity model can clearly identify such generalizations; if so, the productivity of the H98 system will be further enhanced.

More broadly, the productivity model can be used to detect the metrical properties of all morphological processes.[21] In the study presented here, we have assumed that the learner has not fully mastered the derivational morphology of English: indeed, the stress shifting properties of derivational suffixes is acquired quite late, partly having to do with their low frequencies in the linguistic data (Jarmulowicz 2002). Here we sample a few representative derivational suffixes and explore their roles in affecting the stress contour of the stem; some of these, as we shall see, have exceptions and thus pose some challenges to a learning model. For instance, the suffix *-ary* is generally taken to be stress preserving as in *stàtion–stàtionary* but there are also pairs such as

[21]It can be used to detect the productivity of morphological rules/affixes. Some examples are already reviewed in section 2; for a comprehensive treatment, see Yang (in prep).

dòcument-documèntary where the stress does shift. Again using the morpheme segmentations provided in the English Lexicon Project, we compare the stress pattern of the stem and the suffixed form, while omitting words for which stress shift is not applicable (i.e., monosyllabic stems such as *tone-tonic*). For all four suffixes, we consider whether the non-shifted variant is productive, as this is the assumption of the child at the time of acquisition −− the child has learned that suffixes do not shift in English, see the discussion of e.g *-ing* above. The results for stress preserving *ment* and *ary* are are summarized in Table 3. We see that the stress preserving suffix *ary* remains productively so despite a few counterexamples.

suffix	shifting	N	m	valid
-ment	no	201	0	yes
-ary	no	41	8	yes[a]

Table 3: The validity of stress preservation for certain derivational suffixes that are factually stress preserving. a. $8 < 41/\ln 41 = 11$.

As seen in table 4, for the stress shifting suffixes *ic* and *ous*, the non-shifting option is non-productive. The shifting option, in contrast, is exceptionless, assuming that the child analyses *ous* using the stress pattern for nouns.

suffix	shifting	N	m	valid
-ic	no	135	120	no
-ous	no	90	30	no[a]

Table 4: The validity of stress preservation for certain derivational suffixes that are factually stress shifting. a. $30 > 90/\ln 90 = 20$.

6 Conclusion

Given the complexity of the English metrical system and its interactions with the other components of grammar, our treatment here is admittedly preliminary. We do hope, however, that the quantitative approach guided by a precise model of learning can be used to evaluate the theories of metrical stress from the past and shed light on the directions of research in the future. And we hope that this study makes a suitable tribute to Carol Chomsky's legacy:

> The information thus revealed about discrepancies between child grammar and adult grammar affords considerable insight into the process of acquisition, and in addition,

into the nature of the structures themselves. (Carol Chomsky 1967, p2)

Appendix

(lex-, stem-): need $5662/\ln 5662 = 5007$.

- HV87: noun rules consistent with 4819 $\Longrightarrow$ No.
- H97: noun rules consistent with 4906 $\Longrightarrow$ No.

(lex-, stem+): need $4138/\ln 4138 = 3641$.

- HV87: noun rules consistent with 3454 $\Longrightarrow$ No.
- H97: noun rules consistent with 3720 $\Longrightarrow$ Yes.

(lex+, stem-): need $4047/\ln 4047 = 3560$ for nouns, $2291/\ln 2291 = 1995$ for verbs.

- HV87:

  consistent with 3479 nouns $\Longrightarrow$ No.

  consistent with 2052 verbs $\Longrightarrow$ Yes.

- H97:

  consistent with 3771 nouns $\Longrightarrow$ Yes.

  consistent with 2052 verbs $\Longrightarrow$ Yes.

(lex+, stem+): need $3102/\ln 3102 = 2716$ for nouns, $3102/\ln 1036 = 887$ for verbs

- HV87:

  consistent with 2607 nouns $\Longrightarrow$ No.

  consistent with 916 verbs $\Longrightarrow$ Yes.

- H98:

  consistent with 2867 nouns $\Longrightarrow$ Yes.

  consisten with 916 verbs $\Longrightarrow$ Yes.

References

Anderson, Stephen R. 1969. West Scandinavian vowel systems and the ordering of phonological rules. Doctoral dissertation, MIT, Cambridge, MA.

Baker, M. (2001). Atoms of language. New York: Basic Books.

Balota, D.A., Yap, M.J., Cortese, M.J., Hutchison, K.A., Kessler, B., Loftis, B., Neely, J.H., Nelson, D.L., Simpson, G.B., & Treiman, R. (2007). The English Lexicon Project. Behavior Research Methods, 39, 445-459.

Bortfeld, H., Morgan, J., Golinkoff, R., & Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech stream segmentation. *Psychological Science*, 16, 298-304.

Brill, E. (1995). Transformation-based error-driven learning and natural language processing: A case study in part of speech tagging. *Computational Linguistics*, 21, 543-565.

Bromberger, S. & Halle, M. (1989). Why phonology is different. Linguistic Inquiry, 20, 51-70.

Burzio, Luigi. 1994. Principles of English stress. Cambridge: Cambridge University Press.

Chomsky, Carol. (1969). The acquisition of syntax in children from 5 to 10. Cambridge, MA: MIT Press.

Chomsky, Noam, and Morris Halle. 1968. The sound pattern of English. New York: Harper & Row. [Paperback reprint 1993, Cambridge, Mass.: MIT Press.]

Chomsky, Noam. (1965). Aspects of the Theory of Syntax. Cambridge, MA: MIT Press.

Chomsky, Noam. 1955. Logical structures of linguistic theory. Manuscript, Harvard University and MIT. Published in 1975 Plenum.

Clahsen, H. (1999). Lexical entries and rules of language. *Brain and Behavioral Sciences*, 22:991-1013.

Clahsen, H., Hadler, M., & Weyerts, H. (2004). Speeded production of inflected words in children and adults. *Journal of Child Language*, 31:683-712.

Cutler, Anne and David Swinney. 1979. The access and processing of idiomatic expressions. Journal of Verbal Learning and Verbal Behavior. 18, 523-534.

Dresher, E. (1999). Charting the Learning Path: Cues to Parameter Setting. *Linguistic Inquiry*, 30:27-67.

Dresher, E., & Kaye, J. (1990). A Computational Learning Model for Metrical Phonology. *Cognition*, 34:137-195.

Fikkert, Paula. (1994). On the acquisition of prosodic structure. Doctoral dissertation. Leiden University.

Forster, K. (1976). Accessing the mental lexicon. In Wales, R. & Walker, E. (Eds.) *New approaches to language mechanisms*. Amsterdam: North Holland. 257-287.

Forster, K. (1992). Memory-addressing Mechanisms and Lexical Access. In Frost, R. & Katz, L. (Eds.). *Orthography, Phonology, Morphology, and Meaning*. Amsterdam: Elsevier. 413-434.

Golinkoff, Roberta, Hirsh-Pasek, Kathryn, Cauley, Kathleen & Gordon, Laura. (1987). The eyes have it: lexical and syntactic comprehension in a new paradigm. Journal of Child Language. 14, 23-45.

Gorman, K. (2012). Words and Gaps. Doctoral dissertation. Department of Linguistics. University of Pennsylvania. Forthcoming.

Halle, M. (1962). Phonology in generative grammar. Word. 18, 54-72.

Halle, M. (1978). Knowledge unlearned and untaught: What speakers know about the sounds of their language. In M. Halle, J. Bresnan & G. Miller (Eds.) *Linguistic Theory and Psychological Reality*. Cambridge, MA: MIT Press. 294-303

Halle, M. (1998). The stress of English words 1968-1998. *Linguistic Inquiry*, 29, 539-568.

Halle, M. 1973. Prolegomena to a Theory of Word Formation, Linguistic Inquiry 4, 3-16.

Halle, M. & Kenstowicz, M. (1991). The Free Element Condition and cyclic vs. noncyclic stress. Linguistic Inquiry, 22, 457-501.

Halle, M., & Mohanan, K.-P. (1985). Segmental Phonology of Modern English. *Linguistic Inquiry*, 16:57-116.

Halle, Morris, and Alec Marantz. 1993. Distributed Morphology and the pieces of inflection. In The view from Building 20, ed. by Kenneth Hale and Samuel Jay Keyser, 111âĂŞ176. Cambridge, MA: MIT Press.

Halle, Morris, and Jean-Roger Vergnaud. 1987. An essay on stress. Cambridge, Mass.: MIT Press.

Halle, Morris. 1997. Distributed Morphology: Impoverishment and fission. In PF: Papers at the interface, ed. by Benjamin Bruening, Yoonjung Kang, and Martha McGinnis, 425âĂŞ450. MIT Working Papers in Linguistics 30. Cambridge, MA: MIT, Department of Linguistics and Philosophy, MITWPL.

Hayes, Bruce. 1982. Extrametricality and English stress. Linguistic Inquiry, 13, 227-276.

Hayes, Bruce. 1995. Metrical stress theory. Chicago: University of Chicago Press.

Idsardi, W. (1992). The computation of prosody. Doctoral dissertation. Department of Linguistics and Philosophy. Massachusetts Institute of Technology.

Jarmulowicz, Linda. (2002). English derivational suffix frequency and children's stress judgments. Brain and Language. 81, 192-204.

Jusczyk, P., Cutler, A., & Redanz, N. (1993). Preference for the predominant stress patterns of English worlds. *Child Development*, 64, 675-687.

Jusczyk, P., Friederici, A. D., Wessels, J. M. I., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants's sensitivity to the sound patterns of native language words. *Journal of Memory and Language*, 32, 402âĂŞ420.

Jusczyk, P., & Aslin, R. (1995). Infant's detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, 46, 65-97.

Jusczyk, P., & Hohne, E. A. (1997). Infants' memory for spoken words. *Science*, 277, 1984âĂŞ1986.

Kahn, D. (1976). Syllable-based generalizations in English phonology. Doctoral dissertation. Department of Linguistics and Philosophy. Massachusetts Institute of Technology.

Kehoe, Margaret & Stoel-Gammon, Carol. (1997). The acquisition of prosodic structure: An investigation of current accounts of children's prosodic development. Language. 73, 113-144.

Kiparsky, Paul. 1973. "Elsewhere" in phonology. In A festschrift for Morris Halle, ed. by Stephen R. Anderson and Paul Kiparsky, 93âĂŞ106. New York: Holt, Rinehart and Winston.

Kuhl, P., Williams, K., Lacerda, F., Stevens, K., & Lindblom, B. (1992). Linguistic Experience Alters Phonetic Perception in Infants by 6 Months of Age. *Science*, 255:606-608.

Legate, J. A.,& Yang, C. (2007). Morphosyntactic Learning and the Development of Tense. *Language Acquisition*, 14 (3), 315-344.

Liberman, Mark, and Alan S. Prince. 1977. On stress and linguistic rhythm. Linguistic Inquiry 8:249-336.

Marcus, G., Brinkmann, U., Clahsen, H., Wiese, R., & Pinker, S. (1995). German Inflection: the Exception that Proves the Rule. *Cognitive Psychology*, 29:189-256.

Marcus, G., Pinker, S., Ullman, M., Hollander, M., Rosen, J., & Xu, F. (1992). *Overregularization in Language Acquisition*. *Monographs of the Society for Research in Child Development*, No. 57.

Marslen-Wilson, W. D. & Tyler, L. K. (1997). Dissociating types of mental computation. Nature, 387, 592–594.

Murray, W. & Forster, K. (2004). Serial mechanisms in lexical access: The rank hypothesis. *Psychological Review*, 111, 721-756.

Naigles, L. (1990). Children use syntax to learn verb meanings. Journal of Child Language, 17, 357-374.

Penke, M. & Krause, M. (2002). German noun plurals: A challenge to the dual-mechanism model. *Brain and Language*, 81:303-311.

Pinker, S. (1999). *Words and Rules: the Ingredients of Language*. New York, NY: Basic Books.

Poser, William. 1992. Blocking of phrasal constructions by lexical items. In Sag, I. & Szabolcsi, A. (Eds.) Lexical matters. Stanford: CLSI Publications. 111-131.

Rivest, R. (1976). On self-organizing sequential search heuristics. *Communications of the ACM*, 2:63-67.

Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month-olds. *Science*, 274, 1926âĂŞ1928.

Sereno, J. & Jongman, A. (1997). Processing of English inflectional morphology. *Memory and Cognition*, 25, 425-437.

Sonnestuhl, I. & Huth, A. (2002). Processing and Representation of German -n Plurals: A Dual-mechanism Approach. *Brain and Language*, 81, (1-3), 276-290.

Sternberg, S. (1969). Memory-scanning: Mental processes revealed by reaction-time experiments. American Scientist, 57, 4, 421-457.

Straus, K. (2008). Validations of a Probabilistic Model of Language Learning. Ph.D. Dissertation. Department of Mathematics, Northeastern University. Boston, MA.

Tardif, T., Shatz, M. & Naigles, L. (1997). Caregiver speech and children's use of nouns versus verbs: A comparison of English, Italian, and Mandarin. Journal of Child Language, 24, 535-565.

Tesar, B., & Smolensky, P. (2000). *Learnability in Optimality Theory*. Cambridge, MA: MIT Press.

Tyler, A. & Nagy, W. (1989). The acquisition of English derivational morphology. Journal of Memory and Language. 28, 649-667.

Valian, V. (1986). Syntactic categories in the speech of young children. Developmental Psychology, 22, 562-579.

Werker, J. & Tees, R. (1983). Developmental Change across Childhood in the Perception of Non-native Speech Sounds. *Canadian Journal of Psychology*, 37 (2), 278-286.

Wiese, R. (1996). *The phonology of German*. Cambridge: Cambridge University Press.

Wunderlich, D. (1999). German noun plural reconsidered. Manuscript. University of Düsseldorf. Germany.

Xu, F. & Pinker, S. (1995). Weird Past Tense Forms. *Journal of Child Language*, 22:531-556.

Yang, C. (2002). *Knowledge and learning in natural language*. New York: Oxford University Press.

Yang, C. (2004). Universal grammar, statistics, or both. *Trends in Cognitive Sciences*, 8, 451-456.

Yang, C. (2005). On productivity. *Language Variation Yearbook*, 5, 333-370.

Yang, C. (2006). *The Infinite Gift*. New York: Scribner's.

Yang, C. (2010). Three factors in language variation. *Lingua*, 120, 1160-1177.

Yang, C. (in prep). The price of productivity. Manuscript, University of Pennsylvania.

Yang, C. (in prep). Lexical diffusion is over-regularization: the case of noun/verb stress shift in the history of English.