11

Assessing Child and Adult Grammar

JULIE ANNE LEGATE AND CHARLES YANG*

11.1 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 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 central role in the development of phonological theories (N. Chomsky and Halle 1968; Liberman and Prince 1977; Hayes 1982, 1995; Halle and Vergnaud 1987; Halle 1998) yet considerable disagreement remains. The developmental patterns of stress acquisition may contribute to the understanding of grammatical theories 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

* For helpful comments and suggestions, we would like to thank Morris Halle, Kyle Gorman, and the audiences of the 35th Penn Linguistics Colloquium and the Parallel Domains workshop at USC. An extended version of this paper, Legate and Yang (2011), is available from the authors.

have to make decisions on what constitutes the core system of the grammar—see, e.g., *basic* word orders, *default* rules, *unmarked* 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 idiosyncrasies–see, e.g., diacritics, exception marking, or 'lexical listing'—are frequently asserted without a 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; child's the navigation toward the adult grammar might shed light on the choices of linguistic theorizing. Indeed, one might go as far as to identify the failure of dealing with realistic linguistic input, and exceptions in particular, as the source of a long-standing challenge that has been magnified in recent years. As discussed in by N. Chomsky and Halle (1968), the existence of exceptions and other idiosyncratic patterns that run counter to a theory of grammar is unremarkable unless it leads to the development of a theory with higher degrees of generality since exceptions can be always be memorized. But as illustrated most vividly in the so-called past tense debate, there is a slippery slope from '*some* parts of language are memorized exceptions' to '*all* of language are memorized exceptions'. And the temptation grows stronger by the day as long as one fails to produce the principled treatment of exceptions, and it is presently not difficult to find radically lexicalized theories where everything *is* memorized (e.g., Sag 2010).

Linguistics would seem a dreary enterprise if language were no more than a collection of idiosyncrasies. The burden of proof must fall upon those who do wish to uphold a systematic grammar to develop a principled account for exceptions. Our approach here is learning-theoretic, as we try to develop a realistic acquisition model that operates on the type of data that a young English learner might encounter. As far as we know, no formal study of language acquisition has ever considered the full range of linguistic experience. Keeping to the topic of stress acquisition, all current learning models have been 'sanitized' as they only deal with what the researcher regards as the core patterns of language, thereby steering clear of noise, exceptions, and the like. At the same time, one cannot uncritically assume the ready availability of especially informative items in the input (Tesar and Smolensky 2000); the welfare of the child's metrical stress should not be left to chance—needing to hear words such as *Manitoba* or *Winnipesaukee* (Dresher and Kaye 1990; Dresher 1999).

Our 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 faces, and exactly the type of theoretical choices that the linguist 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.

11.2 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, N. Chomsky and Halle 1968, in particular p. 172). 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.

Though many decision metrics are conceivable, the calculus in our analysis is based on real-time processing complexity of linguistic processes to which there are exceptions. 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 do not follow R. We state without further comment the following result:

(1) **Tolerance Principle**: R is productive if and only if

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

The reader is referred to Yang (2005) for the mathematical details of the model. In essence, the empirical motivation comes from psycholinguistic evidence that the number of exceptions (m) contributes to the time complexity of processing, so much so that after m reaches a certain threshold as specified above, it becomes more efficient to list all N items as exceptions, which can be processed in a frequency-sensitive fashion.

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 percent; Sonnenstuhl and Huth 2002): the *-s* rule would have 5 percent coverage and 95 percent 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[1] feminine nouns that have appeared at least once per million in the Mannheim

[1] 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.

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 and Krauss 2002). 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.[2] 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 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 double as many items as 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

[2] 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 and Pinker 1995; Clahsen 1999).

the learner processes more primary linguistic data. We turn to explain these issues in the acquisition of the metrical stress system of English.

11.3 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

(2) 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 and 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 1978) and is thus capable of building syllables from segments which are subsequently used to construct the metrical system.[3] For instance, Dutch and English-learning infants at nine months 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-a-half months (Jusczyk and Aslin 1995). While the role of statistical learning in word segmentation (Saffran, Aslin, and Newport 1996) is not useful as previously thought, universal constraints on lexical stress (Halle and Vergnaud 1987; Yang 2004) and the bootstrapping use of previously segmented words (Jusczyk and 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 seven-and-a-half-month-old English-learning infants perform better at recognizing trochaic than iambic words (Jusczyk, Cutler, and Redanz 1993; Jusczyk, Houston, and 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.

[3] See Gorman (2012) 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.

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 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 and 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 the surface structure is the task of the grammar, the very target of learning.[4] 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 and Kaye 1990; Dresher 1999),[5] 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 of 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 be appropriately characterized as quantity-insensitive (cf.

[4] 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.

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

Kehoe and 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 (Yang 2011).

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 counter-examples 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 generalization 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 11.2 will play a critical role in our approach to metrical stress. The preliminary success of the model reviewed in section 11.2, 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.

Universal Grammar provides a core set of parametric options that delimit a range of possible metrical structures (syllables, 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: as noted earlier, English exhibits distinct stress patterns for nouns and verbs (see 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:

(3) a. If a grammar fails to reach productivity as prescribed by the Tolerance Principle (1), 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.[6]

[6] 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

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:[7]

(4) $G_1 \rightarrow G_2 \rightarrow G_3 \rightarrow \ldots \rightarrow G_T$

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 of rendering G_i unproductive thereby forcing the learner to adopt a different grammar G_{i+1}.[8] 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.

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 (4), 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 11.3. 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.

11.4 The Learning Process

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. Space limitation prevents us from giving the topic even a cursory review. Roughly

English irregular verbs, in contrast to the direct memorization approach in the dual-route morphology literature (Pinker 1999).

7 Strictly speaking, of course, there is no target grammar that 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 (in prep) for an application of the productivity model to the well-known case of noun/verb diatonic stress shift in the history of English.

8 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 and Halle 1989). For additional discussion, see Yang (2010).

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. Major differences between the models arise largely in the treatment of nouns with long vowels in the final syllable. The influential treatment of Halle and Vergnaud (1987) predicts final primary stress, while Halle's later account (1998), based on a different conception of metrical calculation that needn't concern us here, predicts final secondary stress, except in the case of a final long unstressable syllable, which will not bear stress. We will provide a summary of these predictions momentarily; for the moment, let's develop 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. In total, 4.5 million words are used for a total of about 26,700 distinct types. Using a state-of-the-art part-of-speech tagger based on Brill (1995),[9] we 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 inflectionally formed 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.[10] We ignore the prosodic effects on lexical stress in the present study. We assume that syllables containing long vowels (diphthongs and the

[9] Available at <http://gposttl.sourceforge.net/>.

[10] 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.

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 (4) 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 consider only 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 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 11.1.[11]

The distribution in Table 11.1 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.[12] 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

Table 11.1. Stress patterns for words with frequency ≥ 1 in 10,000

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

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

[12] Fikkert provides evidence, noted immediately below, for this limitation. Also compatible with our model would be for the child not to 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.

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 and 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, and 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 are 4047 nouns, 2402 verbs, and 5763 lexically and prosodically distinct words altogether.[13]

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 percent), 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.

[13] 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.

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 percent 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's 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 and Nagy 1989), which may simply be the result of derivational forms being 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 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. Toward the end of this section, we briefly discuss how the child may acquire the stress-shifting properties of derivational suffixes.

We now turn to the placement of primary stress under the Halle and Vergnaud (1987) and Halle (1998) proposals, which are summarized operationally as follows:

(5) The Halle and 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.

(6) 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 (5b).

Table 11.2 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 consider a grammar to be successful only if its rules reach productivity for both nouns and verbs. The raw data can be found in Legate and Yang (2011).

The H98 system under (**lex+**, **stem+**) can successfully identify the stress patterns of English with a tolerable amount of exceptions. It also manages to reach productivity under (**lex+**, **stem**-) though it accumulates more exceptions and is thus disfavored. 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. Upon inspection, most of these end 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 and Prince 1977) that these suffixes are unstressable and are therefore ignored by the rules for stress assignment. Although he does not address

Table 11.2. Evaluation of stress grammars for words with frequency ≥ 1 per million. a. with 515 exceptions. b. with 355 exceptions

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

Table 11.3. The validity of stress preservation for certain derivational suffixes that are factually stress-preserving. a. $8 < 41/\ln 41 = 11$

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

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.[14] 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 are 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 *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. Another motivation for this treatment is due to the fact that young children may not have carried out derivational segmentation; once derivational suffixes are beginning to be acquired, they are initially assumed to be stress-preserving.

The results for stress-preserving-*ment* and-*ary* are are summarized in Table 11.3. We see that the stress-preserving suffix-*ary* remains productively so despite a few counter-examples.

As seen in Table 11.4, 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 analyzes *-ous* using the stress pattern for nouns.

[14] It can be used to detect the productivity of morphological rules/affixes. Some examples are already reviewed in section 11.2; for a comprehensive treatment, see Yang (in prep).

Table 11.4. The validity of stress preservation for certain derivational suffixes that are factually stress-shifting. a. $30 > 90/\ln 90 = 20$

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

11.5 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 1969: 2)