Opinion TRENDS in Cognitive Sciences Vol.8 No.10 October 2004

Full text provided by www.sciencedirect.com

Universal Grammar, statistics or both?

Charles D. Yang

Department of Linguistics and Psychology, Yale University, 370 Temple Street 302, New Haven, CT 06511, USA

Recent demonstrations of statistical learning in infants have reinvigorated the innateness versus learning debate in language acquisition. This article addresses these issues from both computational and developmental perspectives. First, I argue that statistical learning using transitional probabilities cannot reliably segment words when scaled to a realistic setting (e.g. child-directed English). To be successful, it must be constrained by knowledge of phonological structure. Then, turning to the *bona fide* theory of innateness – the Principles and Parameters framework – I argue that a full explanation of children's grammar development must abandon the domain-specific learning model of triggering, in favor of probabilistic learning mechanisms that might be domain-general but nevertheless operate in the domain-specific space of syntactic parameters.

Two facts about language learning are indisputable. First, only a human baby, but not her pet kitten, can learn a language. It is clear, then, that there must be some element in our biology that accounts for this unique ability. Chomsky's Universal Grammar (UG), an innate form of knowledge specific to language, is a concrete theory of what this ability is. This position gains support from formal learning theory [1–3], which sharpens the logical conclusion [4,5] that no (realistically efficient) learning is possible without a priori restrictions on the learning space. Second, it is also clear that no matter how much of a head start the child gains through UG, language *is* learned. Phonology, lexicon and grammar, although governed by universal principles and constraints, do vary from language to language, and they must be learned on the basis of linguistic experience. In other words, it is a truism that *both* endowment and learning contribute to language acquisition, the result of which is an extremely sophisticated body of linguistic knowledge. Consequently, both must be taken into account, explicitly, in a theory of language acquisition [6,7].

Contributions of endowment and learning

Controversies arise when it comes to the relative contributions from innate knowledge and experience-based learning. Some researchers, in particular linguists, approach language acquisition by characterizing the scope and limits of innate principles of Universal Grammar that govern the world's languages. Others, in particular psychologists, tend to emphasize the role of experience and the child's domain-general learning ability. Such a disparity in research agenda stems from the division of labor between endowment and learning: plainly, things that are built in need not be learned, and things that can be garnered from experience need not be built in.

The influential work of Saffran, Aslin and Newport [8] on statistical learning (SL) suggests that children are powerful learners. Very young infants can exploit transitional probabilities between syllables for the task of word segmentation, with only minimal exposure to an artificial language. Subsequent work has demonstrated SL in other domains including artificial grammar, music and vision, as well as SL in other species [9–12]. Therefore, language learning is possible as an alternative or addition to the innate endowment of linguistic knowledge [13].

This article discusses the endowment versus learning debate, with special attention to both formal and developmental issues in child language acquisition. The first part argues that the SL of Saffran *et al.* cannot reliably segment words when scaled to a realistic setting (e.g. child-directed English). Its application and effectiveness must be constrained by knowledge of phonological structure. The second part turns to the *bona fide* theory of UG – the Principles and Parameters (P&P) framework [14,15]. It is argued that an adequate explanation of children's grammar must abandon the domain-specific learning models such as triggering [16,17] in favor of probabilistic learning mechanisms that may well be domain-general.

Statistics with UG

It has been suggested [8,18] that word segmentation from continuous speech might be achieved by using transitional probabilities (TP) between adjacent syllables A and B, where $TP(A \rightarrow B) = P(AB)/P(A)$, where $P(AB)$ = frequency of B following A, and $P(A)$ = total frequency of A. Word boundaries are postulated at 'local minima', where the TP is lower than its neighbors. For example, given sufficient exposure to English, the learner might be able to establish that, in the four-syllable sequence '*prettybaby*', TP(*pre*→*tty*) and TP(*ba*→*by*) are both higher than TP(*tty*→*ba*). Therefore, a word boundary between *pretty* and *baby* is correctly postulated. It is remarkable that 8-month-old infants can in fact extract three-syllable words in the continuous speech of an artificial language from only two minutes of exposure [8]. Let us call this SL model using local minima, SLM.

Statistics does not refute UG

To be effective, a learning algorithm – or any algorithm – must have an appropriate representation of the relevant (learning) data. We therefore need to be cautious about the

Corresponding author: Charles D. Yang (charles.yang@yale.edu).

www.sciencedirect.com 1364-6613/$ - see front matter © 2004 Elsevier Ltd. All rights reserved. doi:10.1016/j.tics.2004.08.006

interpretation of the success of SLM, as Saffran *et al.* themselves stress [19]. If anything, it seems that their results [8] strengthen, rather than weaken, the case for innate linguistic knowledge.

A classic argument for innateness [4,5,20] comes from the fact that syntactic operations are defined over specific types of representations – constituents and phrases – but not over, say, linear strings of words, or numerous other logical possibilities. Although infants seem to keep track of statistical information, any conclusion drawn from such findings must presuppose that children know *what kind* of statistical information to keep track of. After all, an infinite range of statistical correlations exists: for example, What is the probability of a syllable rhyming with the next? What is the probability of two adjacent vowels being both nasal? The fact that infants can use SLM at all entails that, at minimum, they know the relevant unit of information over which correlative statistics is gathered; in this case, it is the syllable, rather than segments, or front vowels, or labial consonants.

A host of questions then arises. First, how do infants know to pay attention to syllables? It is at least plausible that the primacy of syllables as the basic unit of speech is innately available, as suggested by speech perception studies in neonates [21]. Second, where do the syllables come from? Although the experiments of Saffran *et al.* [8] used uniformly consonant–vowel (CV) syllables in an artificial language, real-world languages, including English, make use of a far more diverse range of syllabic types. Third, syllabification of speech is far from trivial, involving both innate knowledge of phonological structures as well as discovering language-specific instantiations [22]. All these problems must be resolved before SLM can take place.

Statistics requires UG

To give a precise evaluation of SLM in a realistic setting, we constructed a series of computational models tested on speech spoken to children ('child-directed English') [23,24] (see Box 1). It must be noted that our evaluation focuses on the SLM model, by far the most influential work in the SL tradition; its success or failure may or may not carry over to other SL models.

The segmentation results using SLM are poor, even assuming that the learner has already syllabified the input perfectly. Precision is 41.6%, and recall is 23.3% (using the definitions in Box 1); that is, over half of the words extracted by the model are not actual words, and close to 80% of actual words are not extracted. This is unsurprising, however. In order for SLM to be usable, a TP at an actual word boundary must be lower than its neighbors. Obviously, this condition cannot be met if the input is a sequence of monosyllabic words, for which a space must be postulated for every syllable; there are no local minima to speak of. Whereas the pseudowords in Saffran *et al.* [8] are uniformly three-syllables long, much of child-directed English consists of sequences of monosyllabic words: on average, a monosyllabic word is followed by another monosyllabic word 85% of time. As long as this is the case, SLM cannot work in principle.

Yet this remarkable ability of infants to use SLM could still be effective for word segmentation. It must be constrained – like any learning algorithm, however powerful – as suggested by formal learning theories [1–3]. Its performance improves dramatically if the learner is equipped with even a small amount of prior knowledge about phonological structures. To be specific, we assume, uncontroversially, that each word can have only one primary stress. (This would not work for a small and closed set of functional words, however.) If the learner knows this, then '*b**i**gb**a**dw**o**lf*' breaks into three words for free. The learner turns to SLM only when stress information fails to establish word boundary; that is, it limits the search for local minima in the window between two syllables that both bear primary stress, for example, between the two **a**'s in the sequence '*l**a**nguage**a**cquisition*'. This is plausible given that 7.5-month-old infants are sensitive to strong/weak patterns (as in *fallen*) of prosody [22]. Once such a structural principle is built in, the stress-delimited SLM can achieve precision of 73.5% and recall of 71.2%, which compare favorably to the best performance previously reported in the literature [26]. (That work, however, uses a computationally prohibitive algorithm that iteratively optimizes the entire lexicon.)

Modeling results complement experimental findings that prosodic information takes priority over statistical information when both are available [27], and are in the same vein as recent work [28] on when and where SL is effective or possible. Again, though, one needs to be cautious about the improved performance, and several unresolved issues need to be addressed by future work (see Box 2). It remains possible that SLM is not used at all in actual word segmentation. Once the one-word/one-stress principle is built in, we can consider a model that uses no SL, hence avoiding the probably considerable computational cost. (We don't know how infants keep track of TPs, but it is certainly non-trivial. English has thousands of syllables; now take the quadratic for the number of pair-wise TPs.) It simply stores previously extracted words in the memory to bootstrap new words. Young children's familiar segmentation errors (e.g. '*I was have*' from *be-have*, '*hiccing up*' from *hicc-up*, '*two dults*',

Box 1. Modeling word segmentation

The learning data consists of a random sample of child-directed English sentences from the CHILDES database [25] The words were then phonetically transcribed using the Carnegie Mellon Pronunciation Dictionary, and were then grouped into syllables. Spaces between words are removed; however, utterance breaks are available to the modeled learner. Altogether, there are 226 178 words, consisting of 263 660 syllables.

Implementing statistical learning using local minima (SLM) [8] is straightforward. Pairwise transitional probabilities are gathered from the training data, which are then used to identify local minima and postulate word boundaries in the on-line processing of syllable sequences. Scoring is done for each utterance and then averaged. Viewed as an information retrieval problem, it is customary [26] to report both precision and recall of the performance.

Precision = No. of correct words/No. of all words extracted by SLM

Recall = No. of words correctly extracted by SLM/No. of actual words

For example, if the target is '*in the park*', and the model conjectures '*inthe park*', then precision is 1/2, and recall is 1/3.

Box 2. Questions for future research

- Can statistical learning be used in the acquisition of language-specific phonotactics, a prerequisite for syllabification and a prelude to word segmentation?
- Given that prosodic constraints are crucial for the success of statistical learning in word segmentation, future work needs to quantify the availability of stress information in spoken corpora.
- Can further experiments, carried over realistic linguistic input, tease apart the multiple strategies used in word segmentation [22]? What are the psychological mechanisms (algorithms) that integrate these strategies?
- It is possible that SLM is more informative for languages where words are predominantly multisyllabic, unlike child-directed English. More specifically, how does word segmentation, statistical or otherwise, work for agglutinative (e.g. Turkish) and polysynthetic (e.g. Mohawk) languages, where the division between words, morphology and syntax is quite different from more clear-cut cases like English?
- How does statistical learning cope with exceptions in the learning data? For example, idioms and register variations in grammar are apparently restricted (and memorized) for individual cases, and do not interfere with the parameter setting in the core grammar.
- How do processing constraints [33] figure in the application of parameter setting?

from *a-dult*) suggest that this process does take place. Moreover, there is evidence that 8-month-old infants can store familiar sounds in memory [29]. And finally, there are plenty of single-word utterances – up to 10% [30] – that give many words for free. ('Free' because of the stress principle: regardless of the length of an utterance, if it contains only one primary stress, it must be a single word and the learner can then, and only then, file it into the memory). The implementation of a purely symbolic learner that recycles known words yields even better performance: precision = 81.5%, recall = 90.1%.

In light of the above discussion, Saffran *et al.* are right to state that 'infants in more natural settings presumably benefit from other types of cues correlated with statistical information' [8]. Modeling results show that, in fact, the infant *must* benefit from other cues, which are probably structural and do not correlate with statistical information. Although there remains the possibility that some other SL mechanisms might resolve the difficulties revealed here – and they would be subject to similar rigorous evaluations – it seems fair to say that the simple UG principles, which are linguistically motivated and developmentally attested, can help SLM by providing specific constraints on its application. On the other hand, related work that augmented generative phonology with statistics to reanalyze the well-known problem of English past tense acquisition has produced novel results [7].

We now turn to another example where SL and UG are jointly responsible for language acquisition: the case of parameter setting.

UG with statistics

In the P&P framework [14,15], the principles are universal and innate, and their presence and effects can be tested in children – ideally as young as possible – through experimental and corpus-based means. This has largely been successful [31,32]. By contrast, the parameter values, specific to each language, must be learned on the basis of linguistic experience.

Triggering versus variational learning

The dominant approach to parameter setting is 'triggering' ([16]; and see [17] for a derivative). According to this model, the learner is, at any given time, identified with a specific parameter setting. Depending on how the current grammar fares with incoming sentences, the learner can modify some parameter value(s) and obtain a new grammar. (See [33–35] for related models, and [36] for a review.)

There are well-known problems with the triggering approach. Formally, convergence to the target grammar generally cannot be guaranteed [37]. But the empirical problems are more serious. First, triggering requires that at all times, the child's grammar be identified with one of the possible grammars in the parametric space. Indeed, Hyams' ground-breaking work attempts to link null subjects in child English ('*Tickles me*') to 'pro-drop' grammars such as Italian [38] or 'topic-drop' grammars such as Chinese [39], for which missing pronoun subjects are perfectly grammatical. However, quantitative comparisons with Italian and Chinese children, whose propensity of pronoun use are close to Italian and Chinese adults, reveal irreconcilable differences [40,41]. Second, if the triggering learner jumps from one grammar to another, then one expects sudden qualitative and quantitative changes in children's syntactic production. This in general is untrue; for example, subject drop disappears only gradually [42]. These problems with triggering models diminish the psychological appeal of the P&P framework, which has been spectacularly successful in cross-linguistic studies [43].

In our opinion, the P&P framework will remain an indispensable component in any theory of language acquisition, but under the condition that the triggering model, a vestige of the Gold paradigm of learning [1], be abandoned in favor of modern learning theories where learning is probabilistic [2,3]. One class of model we have been pursuing builds on the notion of competition among hypotheses (see also [44]). It is related to one of the earliest statistical learning mechanisms ever studied [45,46], which appears to be applicable across various cognitive and behavioral domains. Our adaptation, dubbed the 'variational model' building on insights from biological evolution [47], identifies the hypothesis space with the grammars and parameters defined by innate UG, rather than a limited set of behavioral responses in the classic studies. Schematically, learning goes as follows (see [6,7] for mathematical details.)

For an input sentence, s, the child:

(i) with probability P_i selects a grammar G_i,
(ii) analyzes s with G_i,
(iii) if successful, reward G_i by increasing P_i, otherwise punish G_i by decreasing P_i.

Hence, in this model, grammars compete in a Darwinian fashion. It is clear that the target grammar will necessarily eliminate all other grammars, which are, at best, compatible with only a portion of the input. The variational model has the characteristics of the selectionist approach to learning and growth [48,49], which has

been invoked in other contexts of language acquisition [50,51]. Specific to syntactic development, three desirable consequences follow. First, the rise of the target grammar is gradual, which offers a close fit with language development. Second, non-target grammars stick around for a while before they are eliminated. They will be exercised, probabilistically, and thus lead to errors in child grammar that are 'principled' ones, for they reflect potential human grammars. Finally, the speed with which a parameter value rises to dominance is correlated with how *incompatible* its competitor is with the input – a fitness value, in terms of population genetics. By estimating frequencies of relevant input sentences, we obtain a quantitative assessment of longitudinal trends along various dimensions of syntactic development.

Competing grammars

To demonstrate the reality of competing grammars, we need to return to the classic problem of null subjects. Regarding subject use, languages fall into four groups defined by two parameters [52]. Pro-drop grammars like Italian rely on unambiguous subject–verb agreement morphology, at least as a necessary condition. Topic-drop grammars like Chinese allow the dropping of the subject (and object) as long as they are in the discourse topic. English allows neither option, as indicated by the obligatory use of the expletive subject '*there*' as in '*There is a toy on the floor*'. Languages such as Warlpiri and American Sign Language allow both.

An English child, by hypothesis, has to knock out the Italian and Chinese options (see [7] for further details). The Italian option is ruled out quickly: most English input sentences, with impoverished subject–verb agreement morphology, do not support pro-drop. The Chinese option, however, takes longer to rule out; we estimate that only about 1.2% of child-directed sentences contain an expletive subject. This means that when English children drop subjects (and objects), they are using a Chinese-type topic-drop grammar. Two predictions follow, both of which are strongly confirmed.

First, we find strong parallels in null-subject use between child English and adult Chinese. To this end, we note a revealing asymmetry of null subjects in the topic-drop grammar. In Chinese, when a topic (TOP) is fronted ('topicalized'), subject drop is possible only if TOP does not interfere with the linking between the dropped subject and the established discourse topic. In other words, subject drop is possible when an adjunct is topicalized (as in 1. below), but not when an argument is topicalized (2.). Suppose the discourse topic below is 'John', denoted by *e* as the intended missing subject, and *t* indicates the trace of TOP (in italics):

1. *Mingtian*, [e guiji [t hui xiayu]]. (*e*=John)
   Tomorrow, [e estimate [t will rain]]
   'It is tomorrow that John believes will rain.'

2. **Bill*, [e renwei [t shi jiandie]] (*e*=John)
   Bill, [e believe [t is spy]]
   'It is Bill that John believes is a spy.'

When we examine English children's missing subjects in *Wh*-questions (the equivalent of topicalization), we find the identical asymmetry. For instance, during Adam's null subject stage [25], 95% (114/120) of *Wh*-questions with missing subjects are adjunct questions ('*Where* e going t?'), whereas very few, 2.8% (6/215), of object/argument questions drop subjects ('**Who* e hit t?'). This asymmetry cannot be explained by performance factors but only follow from an explicit appeal to the topic-drop grammatical process.

Another line of evidence is quantitative and makes cross-linguistic comparisons. According to the present view, when English children probabilistically access the Chinese-type grammar, they will also omit objects when facilitated by discourse. (When they access the English-type grammar, there is no subject/object drop.) The relative ratio of null objects and null subjects is predicted to be constant across English and Chinese children of the same age group, the Chinese children showing adult-level percentages of subject and object use (see [7] for why this is so). This prediction is confirmed in Figure 1.

Developmental correlates of parameters

The variational model is among the first to relate input statistics directly to the longitudinal trends of syntactic acquisition. All things being equal, the timing of setting a parameter correlates with the frequency of the necessary evidence in child-directed speech. Table 1 summarizes several cases.

These findings suggest that parameters are more than elegant tools for syntactic descriptions; their psychological status is strengthened by the developmental correlate in children's grammar, that the learner is sensitive to specific types of input evidence relevant for the setting of specific parameters. With variational learning, input *matters*, and its cumulative effect can be directly combined with a theory of UG to explain child language. On the basis of corpus

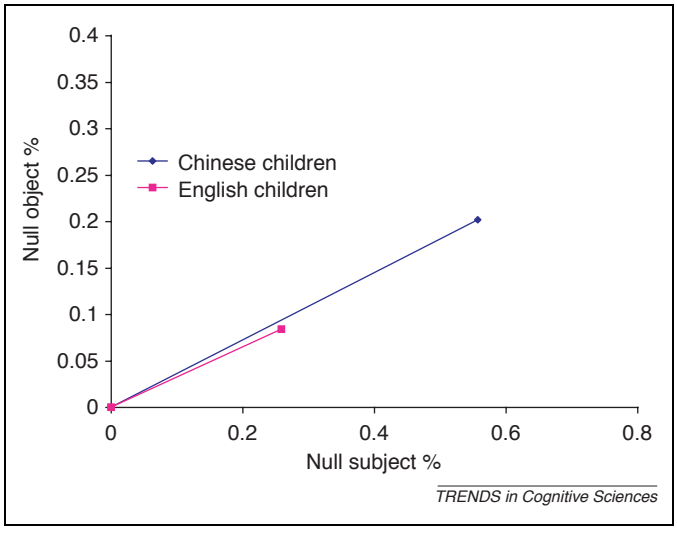

Figure 1. Percentages of null objects and null subjects for English and Chinese children, based on data from [41]. Because English children access the topic-drop grammar with a probability *P* (which gradually reverts to 0), the frequencies of their null subjects and null objects are predicted to be *P* times those of Chinese children. It thus follows the ratio of of null objects to null subjects (the slope of the graph) should be the same for English and Chinese in the same age group, as is the case. Reprinted from [7] with permission from Oxford University Press.

Table 1. Input and output in parameter setting

Parameter	Target language	Requisite evidence	Input (%)	Time of acquisition
Wh fronting[a]	English	*Wh*-questions	25	very early [53]
verb raising[b]	French	verb adverb	7	1;8 [54]
obligatory subject	English	expletive subjects	1.2	3;0 [40,41]
verb second[c]	German/Dutch	OVS sentences [7,35]	1.2	3;0–3;2 [55]
scope marking[d]	English	long-distance *wh*-questions	0.2	4;0+[56]

[a]English moves *Wh*-words in questions; in languages like Chinese, *Wh*-words stay put.
[b]In language like French, the finite verb moves past negation and adverbs ('*Jean voit souvent/pas Marie*'; 'Jean sees often/not Marie'), in contrast to English.
[c]In most Germanic languages, the finite verb takes the second position in the matrix clause, following one and exactly one phrase (of any type).
[d]In German, Hindi and other languages, long-distance *Wh*-questions leave intermediate copies of *Wh*-markers: '*Wer glaubst du wer Recht hat*?'; 'Who think you who right has?' (Who do you think has right?). For children to know that English doesn't use this the option, long-distance *Wh*-questions must be heard in the input. For many children, the German-option persists for quite some time, producing sentences like '*Who do you think who is the box*?' [56].

statistics, this line of thinking can be used to quantify the argument for innateness from the poverty of the stimulus [5,22,57]. In addition, the variational model allows the grammar and parameter probabilities to be values other than 0 and 1 should the input evidence be inconsistent; in other words, two opposite values of a parameter must coexist in a mature speaker. This straightforwardly renders Chomsky's UG compatible with the Labovian studies of continuous variations at both individual and population levels [58]. As a result, the learning model extends to a model of language change [59], which agrees well with the findings in historical linguistics [60] that language change is generally (i) gradual, and (ii) exhibits a mixture of different grammatical options. But these are possible only if one adopts an SL model where parameter setting is probabilistic.

Conclusion

Jerry A. Fodor, one of the strongest advocates of innateness, recently remarked: '…Chomsky can with perfect coherence claim that innate, domain specific PAs [propositional attitudes] mediate language acquisition, while remaining entirely agnostic about the domain specificity of language acquisition *mechanisms*.' ([61], p. 107–8). Quite so. There is evidence that statistical learning, possibly domain-general, is operative at both low-level (word segmentation) as well as high-level (parameter setting) processes of language acquisition. Yet, in both cases, it is constrained by what appears to be innate and domain-specific principles of linguistic structures, which ensure that learning operates on specific aspects of the input; for example, syllables and stress in word segmentation, expletive subject sentences in parameter setting. Language acquisition can then be viewed as a form of 'innately guided learning', where UG instructs the learner 'what cues it should attend to' ([62], p. 85). Both endowment and learning are important to language acquisition – and both linguists and psychologists can take comfort in this synthesis.

Acknowledgements

I am grateful to Tim Gambell, who collaborated on the computational modeling of word segmentation. For helpful comments on a previous draft, I would like to thank Stephen R. Anderson, Norbert Hornstein, Julie Anne Legate, and the anonymous reviewers.

References

1 Gold, E.M. (1967) Language identification in the limit. *Inform. Control* 10, 447–474
2 Valiant, L. (1984) A theory of the learnable. *Commun. ACM* 27, 1134–1142
3 Vapnik, V. (1995) *The Nature of Statistical Learning Theory*, Springer
4 Chomsky, N. (1959) Review of verbal behavior by B.F. Skinner. *Language* 35, 26–57
5 Chomsky, N. (1975) *Reflections on Language*, Pantheon
6 Yang, C.D. (1999) A selectionist theory of language development. In *Proceedings of 37th Meeting of the Association for Computational Linguistics*, pp. 431–435, Association for Computational Linguistics, East Stroudsburg, PA
7 Yang, C.D. (2002) *Knowledge and Learning in Natural Language*, Oxford University Press
8 Saffran, J.R. *et al*. (1996) Statistical learning by 8-month old infants. *Science* 274, 1926–1928
9 Gomez, R.L. and Gerken, L.A. (1999) Artificial grammar learning by one-year-olds leads to specific and abstract knowledge. *Cognition* 70, 109–135
10 Saffran, J.R. *et al*. (1999) Statistical learning of tone sequences by human infants and adults. *Cognition* 70, 27–52
11 Fiser, J. and Aslin, R.N. (2002) Statistical learning of new visual feature combinations by infants. *Proc. Natl. Acad. Sci. U. S. A.* 99, 15822–15826
12 Hauser, M. *et al*. (2001) Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition* 78, B41–B52
13 Bates, E. and Elman, J. (1996) Learning rediscovered. *Science* 274, 1849–1850
14 Chomsky, N. (1981) *Lectures on Government and Binding Theory*, Foris, Dordrecht
15 Chomsky, N. (1995) *The Minimalist Program*, MIT Press
16 Gibson, E. and Wexler, K. (1994) Triggers. *Linguistic Inquiry* 25, 355–407
17 Nowak, M.A. *et al*. (2001) Evolution of universal grammar. *Science* 291, 114–118
18 Chomsky, N. (1955/1975) *The Logical Structure of Linguistic Theory* (orig. manuscript 1955, Harvard University and Massachusetts Institute of Technology), publ. 1975, Plenum
19 Saffran, J.R. *et al*. (1997) Letters. *Science* 276, 1177–1181
20 Crain, S. and Nakayama, M. (1987) Structure dependency in grammar formation. *Language* 63, 522–543
21 Bijeljiac-Babic, R. *et al*. (1993) How do four-day-old infants categorize multisyllabic utterances. *Dev. Psychol.* 29, 711–721
22 Jusczyk, P.W. (1999) How infants begin to extract words from speech. *Trends Cogn. Sci.* 3, 323–328
23 Gambell, T. and Yang, C.D. (2003) Scope and limits of statistical learning in word segmentation. In *Proceedings of the 34th North-eastern Linguistic Society Meeting* (NELS), pp. 29–30, Stony Brook University, New York
24 Gambell, T. and Yang, C.D. Computational models of word segmentation. *Proceedings of the 20th International Conference on Computational Linguistics* (in press)
25 MacWhinney, B. (1995) *The CHILDES Project: Tools for Analyzing Talk*, Erlbaum
26 Brent, M. (1999) Speech segmentation and word discovery: a computational perspective. *Trends Cogn. Sci.* 3, 294–301

27 Johnson, E.K. and Jusczyk, P.W. (2001) Word segmentation by 8-month-olds: when speech cues count more than statistics. *J. Mem. Lang.* 44, 1–20
28 Newport, E.L. and Aslin, R.N. (2004) Learning at a distance: I. Statistical learning of nonadjacent dependencies. *Cogn. Psychol.* 48, 127–162
29 Jusczyk, P.W. and Hohne, E.A. (1997) Infants' memory for spoken words. *Science* 277, 1984–1986
30 Brent, M.R. and Siskind, J.M. (2001) The role of exposure to isolated words in early vocabulary development. *Cognition* 81, B33–B44
31 Crain, S. (1991) Language acquisition in the absence of experience. *Behav. Brain Sci.* 14, 597–650
32 Guasti, M. (2002) *Language Acquisition: The Growth of Grammar*, MIT Press
33 Fodor, J.D. (1998) Unambiguous triggers. *Linguistic Inquiry* 29, 1–36
34 Dresher, E. (1999) Charting the learning path: cues to parameter setting. *Linguistic Inquiry* 30, 27–67
35 Lightfoot, D. (1999) *The Development of Language: Acquisition, Change and Evolution*, Blackwell
36 Yang, C.D. Grammar acquisition via parameter setting. In *Handbook of East Asian Psycholinguistics* (Bates, E. *et al.*, eds), Cambridge University Press. (in press)
37 Berwick, R.C. and Niyogi, P. (1996) Learning from triggers. *Linguistic Inquiry* 27, 605–622
38 Hyams, N. (1986) *Language Acquisition and the Theory of Parameters*, Reidel, Dordrecht
39 Hyams, N. (1991) A reanalysis of null subjects in child language. In *Theoretical Issues in Language Acquisition: Continuity and Change in Development* (Weissenborn, J. *et al*., eds), pp. 249–267, Erlbaum
40 Valian, V. (1991) Syntactic subjects in the early speech of American and Italian children. *Cognition* 40, 21–82
41 Wang, Q. *et al*. (1992) Null subject vs. null object: some evidence from the acquisition of Chinese and English. *Language Acquisition* 2, 221–254
42 Bloom, P. (1993) Grammatical continuity in language development: the case of subjectless sentences. *Linguistic Inquiry* 24, 721–734
43 Baker, M.C. (2001) *Atoms of Language*, Basic Books
44 Roeper, T. (2000) Universal bilingualism. *Bilingual. Lang. Cogn.* 2, 169–186
45 Bush, R. and Mosteller, F. (1951) A mathematical model for simple learning. *Psychol. Rev.* 68, 313–323
46 Atkinson, R. *et al*. (1965) *An Introduction to Mathematical Learning Theory*, Wiley
47 Lewontin, R.C. (1983) The organism as the subject and object of evolution. *Scientia* 118, 65–82
48 Edelman, G. (1987) *Neural Darwinism: the Theory Of Neuronal Group Selection*, Basic Books
49 Piattelli-Palmarini, M. (1989) Evolution, selection and cognition: from 'learning' to parameter setting in biology and in the study of Language. *Cognition* 31, 1–44
50 Mehler, J. *et al*. (2000) How infants acquire language: some preliminary observations. In *Image, Language, Brain* (Marantz, A. *et al*., eds), pp. 51–75, MIT Press
51 Werker, J.F. and Tees, R.C. (1999) Influences on infant speech processing: toward a new synthesis. *Annu. Rev. Psychol.* 50, 509–535
52 Huang, J. (1984) On the distribution and reference of empty pronouns. *Linguistic Inquiry* 15, 531–574
53 Stromswold, K. (1995) The acquisition of subject and object questions. *Language Acquisition* 4, 5–48
54 Pierce, A. (1992) *Language Acquisition and Syntactic Theory: A Comparative Analysis of French and English Child Grammars*, Kluwer
55 Clahsen, H. (1986) Verbal inflections in German child language: acquisition of agreement markings and the functions they encode. *Linguistics* 24, 79–121
56 Thornton, R. and Crain, S. (1994) Successful cyclic movement. In *Language Acquisition Studies in Generative Grammar* (Hoekstra, T. and Schwartz, B. eds), pp. 215–253, Johns Benjamins
57 Legate, J.A. and Yang, C.D. (2002) Empirical reassessment of poverty of stimulus arguments. Reply to Pullum and Scholz. *Linguistic Rev.* 19, 151–162
58 Labov, W. (1969) Contraction, deletion, and inherent variability of the English copula. *Language* 45, 715–762
59 Yang, C.D. (2000) Internal and external forces in language change. *Lang. Variation Change* 12, 231–250
60 Kroch, A. (2001) Syntactic change. In *Handbook of Contemporary Syntactic Theory* (Baltin, M. and Collins, C. eds), pp. 699–729, Blackwell
61 Fodor, J.A. (2001) Doing without *What's Within*: Fiona Cowie's criticism of nativism. *Mind* 110, 99–148
62 Gould, J.L. and Marler, P. (1987) *Learning by Instinct*, Scientific American, pp. 74–85

Important information for personal subscribers

Do you hold a personal subscription to a *Trends* journal? As you know, your personal print subscription includes free online access, previously accessed via BioMedNet. From now on, access to the full-text of your journal will be powered by **Science Direct** and will provide you with unparalleled reliability and functionality. Access will continue to be free; the change will not in any way affect the overall cost of your subscription or your entitlements.

The new online access site offers the convenience and flexibility of managing your journal subscription directly from one place. You will be able to access full-text articles, search, browse, set up an alert or renew your subscription all from one page.

In order to protect your privacy, we will not be automating the transfer of your personal data to the new site. Instead, we will be asking you to visit the site and register directly to claim your online access. This is one-time only and will only take you a few minutes.

Your new free online access offers you:

• Quick search • Basic and advanced search form • Search within search results • Save search • Articles in press • Export citations • E-mail article to a friend • Flexible citation display • Multimedia components • Help files • Issue alerts & search alerts for your journal

http://www.trends.com/claim_online_access.htm