Volume 120, Issue 5, May 2010 ISSN 0024-3841

ELSEVIER

Lingua

International Review of General Linguistics

Special Issue Section
Formalising syntactic variability
Guest Edited by Bill Haddican and Bernadette Plunkett

LINGUA available on ScienceDirect from Volume 1, published 1949

Available online at www.sciencedirect.com
ScienceDirect

Available online at www.sciencedirect.com
ScienceDirect
Lingua 120 (2010) 1160–1177
Lingua
www.elsevier.com/locate/lingua

Three factors in language variation

Charles Yang
Department of Linguistics & Computer Science, University of Pennsylvania, United States

Received 30 May 2008; received in revised form 28 August 2008; accepted 7 April 2009
Available online 29 May 2009

Abstract

Universal Grammar and statistical generalization from linguistic data have almost always been invoked as mutually exclusive means of explaining child language acquisition. This papers show that such segregation is both conceptually unnecessary and empirically flawed. We demonstrate the utility of general learning mechanisms in the acquisition of the core grammatical system through frequency effects in parameter setting, and develop an optimization-based model of productivity with applications to morphology and syntax in the periphery. These findings in child language support the approach to the evolution of language that seeks connections between language and other cognitive systems, in particular the consequence of general principles of efficient computation.

Keywords: Parameter setting; Core vs. periphery; Statistical learning; Morphological processing; Productivity; Dative alternation

1. Introduction

How much should we ask of Universal Grammar? Not too little, for there must be a place for our unique ability to acquire a language along with its intricacies and curiosities. But asking for too much won't do either. A theory of Universal Grammar is a statement of human biology, and one needs to be mindful of the limited structural modification that would have been plausible under the extremely brief history of *Homo sapiens* evolution.

In a recent article, Chomsky (2005: 6) outlines three factors that determine the properties of the human language faculty:

(1) a. Genetic endowment, ''which interprets part of the environment as linguistic experience . . . and which determines the general course of the development of the language faculty''
b. Experience, ''which leads to variation, within a fairly narrow range, as in the case of other subsystems of the human capacity and the organism generally''.
c. Principles not specific to the faculty of language: ''(a) principles of data analysis that might be used in language acquisition and other domains; (b) principles of structural architecture and developmental constraints. . . including principles of efficient computation''

These factors have been frequently invoked to account for linguistic variation—in almost always mutually exclusive ways, perhaps for the understandable reason that innate things needn't be learned and vice versa. Rather than

E-mail address: charles.yang@ling.upenn.edu.

0024-3841/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.lingua.2008.09.015

dwelling on these efforts – see Yang (2004) for an assessment – we approach the problem of variation from the angle of acquisition by developing a framework in which all three factors are given a fair billing. The study of child language points to two fairly distinct types of language variation, which appear to invoke two distinct mechanisms of language acquisition.

One kind of variation derives from the innate and invariant system of Universal Grammar (1a). Such a space of variation constitutes the initial state of linguistic knowledge, which traditionally has been considered the ‘‘core’’ linguistic system (Chomsky, 1981). The child’s task is one of **selection** from a narrow range of options (e.g., parameter values, constraint rankings) that are realized in her linguistic environment. A prominent line of evidence for the genetic endowment of language comes from the fixed range of linguistic options, some of which are not present in the input data, but which the child nevertheless spontaneously accesses and gradually eliminates during the course of acquisition.

Quite a different type of variation consists of language specific generalizations which are derived from the linguistic environment, i.e., experience (1b). This type of variation can be identified with the periphery of the language faculty (Chomsky, 1981:8): ‘‘marked elements and constructions’’, including ‘‘borrowing, historical residues, inventions’’ and other idiosyncrasies. The child’s task, as we shall see, is one of **evaluation**: decision making processes that determine the scope of inductive generalizations based on the input yet still ‘‘within a fairly narrow range’’. We further suggest that the instantiation of language variation by the child learner follows at least certain principles not specific to the faculty of language (1c). The mechanism which selects amongst alternatives in the core parameter system in (1b) is probabilistic in nature and apparently operates in other cognitive and perceptual systems, and had indeed first been proposed in the study of animal learning and behavior. The acquisition of the periphery system in (1b) reflects general principles of efficient computation which manipulate linguistic structures so as to optimize the time course of online processing, very much in the spirit of the evaluation measure in the earlier studies of generative grammar (Chomsky, 1965; Chomsky and Halle, 1968). Both types of learning mechanisms show sensitivity to certain statistical properties of the linguistic data that have been largely ignored in works that ask too much of Universal Grammar but would be difficult to capture under approaches that rely solely on experience.

We take up these matters in turn.

2. Variation and selection

2.1. Return of the parameter

Saddled with the dual goals of descriptive and explanatory adequacy, the theory of grammar is primed to offer solutions to the problem of language variation and acquisition in a single package. This vision is clearly illustrated by the notion of syntactic parameters (Chomsky, 1981). Parameters unify regularities from (distant) aspects of the grammar both within and across languages, thereby acting as a data compression device that reduces the space of grammatical hypotheses during learning. The conception of parameters as triggers, and parameter setting as flipping switches offers a most direct solution to language acquisition.

There was a time when parameters featured in child language as prominently as in comparative studies. Nina Hyams’ (1986) ground breaking work was the first major effort to directly apply the parameter theory of variation to the problem of acquisition. In recent years, however, parameters have been relegated to the background. The retreat is predictable when broad claims are made that children and adults share the identical grammatical system (Pinker, 1984) or that linguistic parameters are set very early (Wexler, 1998). Even if we accepted these broad assertions, a responsible account of acquisition would still require the articulation of a learning process: a child born in Beijing will acquire a different grammatical system or parameter setting from a child born in New York City, and it would be nice to know how that happens. Unfortunately, influential models of parameter setting (e.g., Gibson and Wexler, 1994, but see Sakas and Fodor, 2001) have failed to deliver formal results (Berwick and Niyogi, 1996),[1] and it has been difficult to bridge the empirical gap between child language and specific parameter settings in the UG space (Bloom, 1993; Valian, 1991; Wang et al., 1992; Yang, 2002). The explanation of child language, which **does** differ from adult language, falls upon either performance limitations or discontinuities in the grammatical system, both of which presumably mature with age and general cognitive development—no thanks to parameters.

[1] Baker (2002) and Snyder (2007) both sketched out properties of the parameter space that would make learning more efficient but no specific learning model has been given.

To return, parameters must provide remedy for both the formal and the empirical problem in child language. The hope, on our view, lies in paying attention to the factors of experience (1b) and the process of learning (1c), which have not been addressed with sufficient clarity in the generative approach to acquisition. The **variational** learning model (Yang, 2002) is an attempt to provide quantitative connections between the linguistic data and the child's grammatical development through the use of parameters. To capture the gradualness of syntactic acquisition, we introduce a probabilistic component to parameter learning, which is schematically illustrated as follows[2];

(2) For an input sentence s, the child
  a. with probability P_i selects a grammar G_i,
  b. analyzes s with G_i
  c. • if successful, reward G_i by increasing P_i
     • otherwise punish G_i by decreasing P_i

Learning the target grammar involves the process of **selection** which eliminates grammatical hypotheses not attested in the linguistic environment; indeed, the variational model was inspired by the dynamics of Natural Selection in biological systems (Lewontin, 1983). It is obvious that non-target grammars, which all have non-zero probabilities of failing in the target grammar environment, will eventually be driven to extinction. The probabilistic nature of learning allows for alternative grammars – more precisely, parameter values – to co-exist, while the target grammar gradually rises to dominance over time.[3] The reality of co-existing grammars has been discussed elsewhere (Yang, 2002, 2006; Legate and Yang, 2007; Roeper, 2000 and subsequent work) but that line of evidence clearly rests on establishing the fact that parameter setting is not **too** early, at least not in all cases; if the child is already on target, the appeal to non-target parameter values as an explanatory device would be vacuous. This, then, requires us to develop a framework in which the time course of parameter setting can be quantitatively assessed.

It can be observed that the rise of a grammar under (2) is a function of the probability with which it succeeds under a sample of the input, as well as that of failure by its competitors. The dynamics of learning can be formalized much like the dynamics of selection in the evolutionary system. Specifically, we can quantify the ''fitness'' of a grammar from the UG Grammar pool as a probability of its failure in a specific linguistic environment:

(3) The *penalty probability* of grammar G_i in a linguistic environment E is[4]

$$c_i = \Pr(G_i \not\to \mathbf{s} | \mathbf{s} \in \mathbf{E})$$

In the idealized case, the target grammar has zero probability of failing but all other grammars have positive penalty probabilities. Given a sufficient sample of the linguistic data, we can estimate the penalty probabilities of the grammars in competition. Note that such tasks are carried out by the scientist, rather than the learner; these estimates are used to quantify the development of parameter setting but do not require the tabulation of statistical information by the child. The present case is very much like the measure of fitness of fruit flies by, say, estimating the probability of them producing viable offspring in a laboratory: the fly does not count anything.[5] Consider two grammars, or two alternative values of a parameter: target G_1 and the competitor G_2, with $c_1 = 0$ and $c_2 > 0$. At any time, $p_1 + p_2 = 1$. When G_1 is selected, p_1 of course will always increase. But when G_2 is selected, p_2 may increase if the incoming data is ambiguous between G_1 and G_2 but it must decrease – with p_1 increasing – when unambiguously G_1 data is presented, an event that occurs with the probability of c_2. That is, the rise of the target grammar, as measured by p_1

[2] The formal details of the learning model can be found in Yang (2002). We will use the terms of ''grammars'' and ''parameters'' interchangeably to denote the space of possible grammars under UG. For analytic results of learnability in a parametric space, see Straus (2008).

[3] The appeal to non-target and linguistically possible options to explain child language can be traced back to Jakobson (1941/1968) and more recently (Roeper, 2000; Crain and Pietroski, 2002; Rizzi, 2004, etc.) though these approaches do not provide an explicit role for either linguistic data or mechanisms of learning.

[4] We write $s \in E$ to indicate that s is an utterance in the environment E, and $G \to s$ to mean that G can successfully analyze s. Formally, the success of $G \to s$ can be defined in any suitable way, possibly even including extra-grammatical factors; a narrow definition that we have been using is simply parsability.

[5] In this sense, the use of the probabilistic information here is distinct from statistical learning models such as Saffran et al. (1996), where linguistic hypotheses **themselves** are derived from the statistical properties of the input data by the learner. See Yang (2004) for an empirical evaluation of that approach.

going to 1, is correlated with the penalty probabilities of its competitor, i.e., c_2, which in turn determines the time course of parameter setting. We turn to these predictions presently.

2.2. *Frequency and parameter setting*

As a starting point, consider the acquisition of verb raising to tense in French and similar languages. First, what is the crucial linguistic evidence that drives the French child to the [+] of the verb raising parameter? Word order evidence such as (4a), where the position of the finite verb is ambiguous, is compatible with both the [+] and [−] value of the parameter, and thus has no effect on grammar selection. Only data of the type in (4b) can unambiguously drive the learner toward the [+] value.

(4) a. Jean voit Marie.
Jean sees Marie.

b. Jean voit souvent/pas Marie.
Jean sees often/not Marie.

The raising of finite verbs in French and similar languages is a very early acquisition. Pierce (1992) reports that in child French as early as 1;8, virtually all verbs preceding *pas* are finite while virtually all verbs following *pas* are non-finite. Bear in mind that children in Pierce's study are still at the two word stage of syntactic development, which is the earliest stage in which verb raising could be observed from naturalistic production. And this early acquisition is due to the accumulative effects of utterances such as (4b), which amount to an estimated 7% of child directed French sentences.[6] Thus we obtain an empirical benchmark for early parameter setting, that 7% of unambiguous input data is sufficient.[7]

If all parameters are manifested at least as frequently as 7% of the input, then parameter setting would indeed be early as widely believed. Fortunately that is not case, for otherwise we would not be able to observe parameter setting in action. Let us consider two major cases of syntactic learning: the Verb Second parameter in languages such as German and Dutch and the obligatory use of grammatical subjects in English. Both have been claimed – incorrectly, as we shall see – to be very early acquisitions on a par with the raising of finite verbs in French (Wexler, 1998).

In an influential paper, Poeppel and Wexler (1993) make the claim that syntactic parameter setting takes place very early, a claim which partially represents an agreement between the competence-based and the performance-based approach (Pinker, 1984; Valian, 1991; Bloom, 1990; Gerken, 1991; cf. Hyams and Wexler, 1993) to grammar acquisition: both sides now consider the child's grammatical system to be adult like. Poeppel & Wexler's study is based on the acquisition of the V2 parameter. They find that in child German, finite verbs overwhelmingly appear in the second (and not final) position while non-finite verbs overwhelmingly appear in the final (and not second) position.

But this does not warrant their conclusion that the V2 parameter has been set. A finite verb in the second position does not mean it has moved to the "V2" position, particularly if the pre-verbal position is filled with a subject, as the examples from Poeppel and Wexler (1993:3–4) illustrate below:

(5) a. Ich hab ein dossen Ball.
I have a big ball

b. Ich mach das nich.
I do that not

The structural position of the verb here deserves additional consideration. It is entirely possible that the verb has gone only as far as T, and the subject would be situated in the Spec of T, and the clausal structure of the raised verb (5) is not like German but like French. The evidence for V2 can be established only when the verb is unambiguously high (e.g., higher than T) and the preverbal position is filled.

[6] Unless noted through the citation of other references, the frequencies of specific linguistic input from child directed data are obtained from the CHILDES database (MacWhinney, 1995). The details can be found in Yang (2002), which was the first generative study of language acquisition that ever used input frequencies in the explanation of child language.

[7] We claim 7% to be sufficient but it may not be necessary; an even lower amount may be adequate if the raising of finite verbs is established before the two word stage, which could be confirmed by comprehension studies such as the preferential looking procedure (Golinkoff et al., 1987).

Table 1
Longitudinal V1 and V2 patterns. All sentences are finite, and the subjects are post-verbal.

Age	V1 sentences	All sentences	V1%
2;4	72	170	43%
2;5	66	132	50%
2;6	147	411	36%
2;7	93	201	46%
2;8	94	292	32%
2;9	98	269	36%
2;10	93	321	28%
2;11	36	259	14%
3;0	56	246	22%
3;1	101	268	37%

To evaluate the setting of the V2 parameter, we must examine finite matrix sentences where the subject is **post-verbal**. In child German acquisition, as shown in the quantitative study of Stromwold and Zimmerman (1999), the subject is consistently placed out of the VP shell and is thus no lower than the specifier position of TP. If so, then a finite verb preceding the subject will presumably by in C, or at least in some node higher than T. Now, if the preverbal, and thus sentence-initial, position is consistently filled, then we are entitled to claim the early setting of the V2 parameter—or however this property comes to be analyzed. But Poeppel and Wexler's claim is not supported: the preverbal position is filled at child language, as shown in Table 1 which is based on Haegeman's (1995: Tables 5 and 6) longitudinal study of a Dutch child's declarative sentences:

We can see that in the earliest stages, there are close to 50% of V1 utterances, in co-existence with V2 patterns, the latter of which gradually increase in frequency.[8] The claim of early V2 setting is therefore not supported.[9] As argued by Lightfoot (1999), Yang (2002) on independent grounds, the necessary evidence for V2 comes from utterances with the pre-verbal position occupied by the object; such data only comes at the frequency of 1% in child-directed speech, which results in a relatively late acquisition at the 36–38th month (Clahsen, 1986). Now we have established an empirical benchmark for the relatively late setting of a parameter.

The quantitative aspects of parameter setting – specifically, the early and late benchmarks – can be further illustrated by differential development of a single parameter across languages. This leads us to the phenomenon of subject drop by English-learning children, one of the most researched topics in the entire history of language acquisition. Prior to 3;0 (Valian, 1991), children learning English leave out a significant number of subjects, and also a small but not insignificant number of objects. However, children learning pro-drop grammars such as Italian and topic-drop grammars such as Chinese are much closer to adult usage frequency from early on. For instance, Valian (1991) reports that Italian children between 2;0–3;0 omit subjects about 70% of the time, which is also the rate of pro-drop by Italian adults reported by Bates (1976) among others. In Wang et al.'s (1992) comparative study of Chinese and American children, they find that 2 year old American children drop subjects at a frequency of just under 30%,[10] which is significantly lower than Chinese children of the same age group—and obviously significantly higher than English speaking adults. By contrast, the difference in subject usage frequency between Chinese children and Chinese adults is not statistically significant.

If the claim of early parameter setting is to be maintained, and that certainly would be fine for Italian and Chinese children, the disparity between adult and child English must be accounted for by non-parametric factors, presumably by either competence or performance deficiencies. Without pursuing the empirical issues of these alternatives, both approaches amount to postulating significant cognitive differences, linguistic or otherwise, between the learners

[8] The data from 3;1 is probably a sampling oddity: all other months are represented by 5 to 10 recording sessions, but 3;1 had only one.

[9] Poeppel and Wexler's work does show, however, that finite verbs raise to a high position (out of the VP shell), and non-finite verbs stay in the base position, and that the child grammar has an elaborate system of functional projections, and thus replicating Pierce's (1992) findings in French acquisition reviewed earlier. Furthermore, we have no quarrel with their more general claim that the child has access to the full grammatical apparatus including functional projections. Indeed, even the V1 patterns displayed in Table 1 demonstrate that the structure of the CP is available to the learner.

[10] 26.8% to be precise. The literature contains some discrepancies in the rate of subject omission. The criteria used by Wang et al. (1992) seem most appropriate as they excluded subject omissions that would have been acceptable for adult English speakers. Following a similar counting procedure but working with different data sets, Phillips (1995) produced similar estimates of subject drop by children.

acquiring different languages: that is, English learning children are more susceptible to competence and/or performance limitations. This of course cannot be ruled out a priori, but requires independent justification.

The variational learning model provides a different view, in which a single model of learning provides direct accounts for the cross-linguistic findings of language development. The time course of parameter setting needn't be uniform: Italian and Chinese children do set the parameter correctly early on but English children take longer. And the reason for such differences is due to the amount of data necessary for the setting of the parameter, which could differ across languages. The following data can unambiguously differentiate the grammars for the Chinese, Italian and English learning children, and their frequencies in the child-directed speech are given as well.

(6) a. [+topic drop, − pro drop] (Chinese): Null objects (11.6%; Wang et al., 1992: Appendix B)
b. [− topic drop, +pro drop] (Italian): Null subjects in object wh-questions (10%)
c. [− topic drop, − pro drop] (English): expletive subjects (1.2%)

The reasoning behind (6) is as follows. For the Chinese type grammar, subject omission is a subcase of the more general process of topic drop, which includes object omission as well. Neither the Italian nor the Chinese type grammar allows that, and hence object omission is the unambiguous indication of the [+] value of the topic drop parameter.[11]

However, topic drop is not without restrictions; such restrictions turn out to differentiate the Chinese type grammar from the Italian type.[12] There is a revealing asymmetry in the use of null subjects in topic drop languages (Yang, 2002) which has not received much theoretical consideration. When topicalization takes place in Chinese, subject drop is possible only if the topic does not interfere with the linking between the null subject and the established discourse topic. In other words, subject drop is possible when an adjunct is topicalized (7a), but not when an argument is topicalized (7b). Suppose the old discourse topic is "John", denoted by *e* as the intended missing subject, whereas the new topic is in italics, having moved from its base position indicated by *t*.

(7) a. *Mingtian*, [*e* guiji [*t* hui xiayu]]. (*e*=John)
Tomorrow, [*e* estimate [*t* will rain]]
"It is tomorrow that John believes it will rain".
b. **Bill*, [*e* renwei [*t*] shi jiandie]]. (*e*=John)
Bill, [*e* believe [*t* is spy]]
"It is Bill that John believes is a spy".

The Itaian pro-drop grammar does not have such restrictions. Following Chomsky (1977) and much subsequent work on topicalization and Wh-movement, the counterpart of (7b) can be identified with an object wh-question. In Italian, subject omission is unrestricted as its licensing condition is through agreement and thus has nothing to do with the discourse and information structure. Here again *e* stands for the omitted subject whereas *t* marks the trace of movement.

(8) a. Chi_t *e* ha baciato *t*?
Who_t *e* has(3SGM) kissed *t*?
'Who has he kissed?'
b. Chi_t e_1 credi che e_2 ami *t*?
Who_t e_1 think(2SG) that e_2 loves(3SGF) *t*?
'Whom do you think she loves?'
c. Dove_t hai *e* visto Maria *t*?
Where_t have(2SG) *e* seen Maria *t*?
'Where have you seen Maria?'

[11] The actual amount of data may be higher: even subject drop would be evidence for the Chinese type grammar since the lack of agreement is actually inconsistent with the pro-drop grammar.

[12] As is well known, the pro drop grammar licenses subject omission via verbal morphology. But "rich" morphology, however it is defined, appears to be a necessary though not sufficient condition, as indicated by the case of Icelandic, a language with "rich" morphology yet obligatory subject. Thus, the mastery of verbal morphology, which Italian and Spanish learning children typically excel at from early on (Guasti, 2002), is not sufficient for the positive setting of the pro-drop parameter.

Upon encountering data such as (8), the Chinese type grammar, if selected by the Italian learning child, would fail and decrease its probability, whose net effect is to increase the probability of the [+] value of the pro drop parameter.

Note that for both the Chinese and Italian learning child, the amount of unambiguous data, namely (6a) and (6b), occur at frequencies great than 7%, the benchmark established from the raising of finite verbs in French. We thus account for the early acquisition of subjects in Italian and Chinese children in Valian (1991), Wang et al. (1992)'s studies.

The English learning child takes longer as the Chinese type grammar lingers on.[13] The rise of the obligatory subject is facilitated by expletive subjects, but such data appear in low frequency: about 1.2% of child-directed English utterances. Using the 1% benchmark established on the relatively late acquisition of the V2 parameter (3;0–3;2; Clahsen, 1986), we expect English children to move out of the subject drop stage roughly around the same time, which is indeed the case (Valian, 1991). And it is not a coincidence that the subject drop stage ends approximately at the same time as the successful learning of the V2 parameter.

Finally, the variational model puts the parameter back into explanations of child language: learner's deviation from the target form is directly explained through parameters, which are also points of variation across languages. The child's language is a statistical ensemble of target and non-target grammars: deviation from the target, then, may bear trademarks of possible human grammars used continents or millennia away, with which the child cannot have direct contact. In the case of null subjects, we can analyze the English learning child as probabilistically using the English type grammar, under which the grammatical subject is always used, in alternation with the Chinese type grammar, under which subject drop is possible if facilitated by discourse. Thus, the distributional patterns of English child null subjects ought to mirror those of Chinese adult null subjects as shown in (7). Indeed, the asymmetry of subject drop under argument/adjunct topicalization for adult Chinese speakers is almost categorically replicated in child English, as summarized below from Adam's subject drop stage (file 1–20; Brown, 1973):

(9) a. 95% (114/120) of Wh-questions with null subjects are adjunct (*how, where*) questions (e.g., ''Where *e* go?'', ''Why *e* working?)
b. 97.2% (209/215) of object questions (*who, what*) contain subjects (e.g., ''What *e* doing?'')

Taken together, we have uncovered significant frequency effects in parameter setting. The fact that frequency plays *some* role in language learning ought to be a truism; language learning is impressively rapid but it does take time. Yet the admission of frequency effects, which can only come about through the admission that experience and learning matter, does not dismiss the importance of the first factor of Universal Grammar (cf. Tomasello, 2003). Quite to the contrary, frequency effects in parameter setting presented here actually strengthen the argument for Universal Grammar; frequency effects are effects about specific linguistic structures. The cases of null subjects and verb second are illustrative because the input data is highly consistent with the target form yet children's errors persist for an extended period of time. To account for such input-output disparities, then, would require the learner to process linguistic data in ways that are quite distant from surface level descriptions. If the space of possible grammars were something like a phrase structure grammar with rules such as ''$S \xrightarrow{\alpha} NP\ VP$'' and ''$S \xrightarrow{\beta} VP$'' where $\alpha + \beta = 1$, it is difficult to see how the phenomenon of subject drop is possible with the vast majority of English sentences containing grammatical subjects. However, if the learner were to approach ''$S \xrightarrow{\alpha}$ [+topic drop]'' and ''$S \xrightarrow{\beta}$ [−topic drop]'', as proposed in the parameter theory, we can capture the empirical findings of children's null subjects—and null objects as well. Parameters have developmental correlates, but they would only turn up when both the input and the learning process are taken seriously.

3. Variation and evaluation

Selection among a universal set of options is by no means the only mode of language acquisition, and it would be folly to attribute all variation in child language to the genetic component of Universal Grammar.

First, and most simply, the size of the search space and the resulting learning time to convergence can increase exponentially with the number of parameters; this may undermine the original conception of parameters as a solution

[13] The Italian type grammar can be swiftly dismissed by the impoverished English morphology, since sufficient agreement is a necessary condition for pro-drop; see Yang (2002: Chapter 4) for additional discussion.

for the problem of explanatory adequacy. One phenomenon, one parameter is not recommended practice for syntax and would not be a wise move for language acquisition either.

Second, it seems highly unlikely that all possibilities of language variation are innately specified; certainly, the acquisition of particular languages does not always exhibit patterns of competition and selection. Variation in the sound system is most obvious. While the development of early speech perception shows characteristics of selectionist learning of phonetic and phonological primitives (Werker and Tees, 1983; Kuhl et al., 1992, see Yang, 2006 for overview), the specific content of morphology and phonology at any point can be highly unpredictable, partly due to ebbs and flows of language change over time; see Bromberger and Halle (1989). Innate principles or constraints packaged in UG notwithstanding, even the most enthusiastic nativist would hesitate to suggest that the English specific rule for past tense formation (''-d'') is one of the options, along with, say, the *é* suffix as in the case of French, waiting to be selected by the child learner. Indeed, the past tense acquisition appears to have an Eureka moment: when the child suddenly comes to the productive use of the ''add -d'' rule, over-regularization of irregular verbs (e.g., *hold-holded*) starts to take place (Marcus et al., 1992; Pinker, 1999).

Close examination of syntactic acquisition reveals that the child is not only drifting smoothly in the land of parameters (section 2) but also taking an occasional great leap forward. A clear example comes from the acquisition of dative constructions. Quantitative analysis of children's speech (Gropen et al., 1989; Snyder and Stromswold, 1997) has shown that not all constructions are learned alike, or at the same time. For 11 of the 12 children in Snyder & Stromsworld's study, the acquisition of double object construction (''I give John a book'') proceeds that of prepositional *to*-construction (''I give a book to John'') by an average of just over 4 months. Prior to that point, children simply reproduce instances of datives present in adult speech. When three-year-olds productively apply these constructions to novel lexical items as in *pilked the cup to Petey* into *I pilked Petey the cup* (Conwell and Demuth, 2007), they must have learned the alternation on the more mundane pairings of *give*, *lend*, *send*, and others.

The acquisition of past tense and datives are obviously different: the specific form of the ''add -d'' rule is learned directly from data whereas the candidate verbs for dative constructions are probably provided by innate and universal syntactic and semantic constraints (Pesetsky, 1995; Harley, 2002; Hale and Keyser, 2002; Rappaport Hovav and Levin, 2008). But the logical problems faced by the learner are the same. Upon seeing a sample that exemplifies a construction or a rule, which may contain exceptions (e.g., irregulars), the learner has to decide whether the observed regularity is a true generalization that extends beyond experience, or one of lexical exceptions that must be stored in memory. For the cases at hand, the answer is positive, but the same decision making process ought to return a negative answer for the rule ''add -t & Rime → ɔ'' when presented with *bring*, *buy*, *catch*, *seek* and *think*. In other words, the decision making involves the recognition of the **productivity** of the language particular processes.

In the rest of this section, we will extend a mathematical model (Yang, 2005) that specifies the conditions under which a rule becomes productive. Even though the empirical motivation for that model is based on morphological learning and processing, there is suggestive evidence that it can extend to the study of syntax as well, as we set out to explain **why** the acquisition of double objects precedes that of prepositional (*to*) dative construction.

3.1. Optimization and productivity

Consider how an English child might learn the past tense rules in her morphology. Suppose she knows only two words *ring-rang* and *sing-sang;* at this point she might be tempted to conjecture a rule ''i→a/__ ŋ'' .[14] The child has every reason to believe this rule to be productive because, at this particular moment, it is completely consistent with the learning data. However, as her vocabulary grows, ''i→a/__ ŋ'' will run into more and more exceptions (e.g, *bring-brought*, *sting-stung*, *swing-swung'*, etc.). Now the learner may decide enough is enough, and the rule will be demoted to the non-productive status: *sing* and *sang* would be memorized as instances of lexical exceptions, which is how English irregular verbs are treated in morphophonology (Halle and Mohanan, 1985; Pinker, 1999). By contrast, the exceptions to the ''add -d'' rule – about 150 irregular verbs, depending how you count – are apparently not enough to derail its productivity, which is backed up by thousands of regular verbs in English language.

[14] One way to do so is to make conservative generalizations over a set of structural descriptions that share the same process of structural change. This is commonly found in inductive learning models in Artificial Intelligence (Mitchell, 1982; Sussman and Yip, 1997), and its first linguistic application is in Chomsky (1955). How the learner comes to such generalizations is an important issue but one which is of no significant interest to us for present purposes; our key task is to determine the productivity of such generalizations.

The question is: How much is enough? How many exceptions can a productive rule tolerate? What kind of batting average[15] would the learner hold out for productive rules? Before we lay out our approach to the problem, let us note that the child learner is superbly adept at recognizing productive processes. For instance, it is well known that English learners over-*regularize* in past tense acquisition (Marcus et al., 1992; Pinker, 1999; Yang, 2002), in up to 10% of all past tense uses. It is perhaps less well known that children do not over-*irregularize*: errors such as *bring-brang* are exceedingly rare, only constituting in 0.2% of the past tense production data (Xu and Pinker, 1995). In Berko's classic Wug study (1958), four year olds reliably supply regular past tense for novel verbs but only one of 86 children extended *gling* and *bing* to the irregular form of *glang* and *bang*, despite maximum similarity to the *sing-sang* and *ring-rang* irregulars. Children generalize productive rules but do not generalize lexical rules.

There is a line of research in morphology that gathers corpus statistics to correlate with behavioral tests for productivity (Aronoff, 1976; Baayen and Lieber, 1991; Hay, 2003 though see Schütze, 2005a,b for a criticism of some of the methodologies) but these results are, at best, a statistical summary of the empirical data that ''accord nicely with [linguists'] intuitive estimate of productivity'' (Baayen and Lieber, 1991). Even if a criterion for productivity established from these data summarizations turns out to be true – e.g., a productive rule can tolerate not more than 25% of exceptions – we still need an explanation why that magic number is 25%, rather than 18% or 32%.

Our approach is a throwback to the notion of an *evaluation measure,* which dates back to the foundations of generative grammar (Chomsky, 1955; 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 warranted or not. It is useful to recall that the evaluation measure ''is not given a priori . . . Rather, a proposal concerning such a measure is an empirical hypothesis about the nature of language'' (Chomsky, 1965:37). A model of productivity, therefore, requires independent motivation. And this is an area where Chomsky's third factor – in particular, ''principles of efficient computation'' – may play an important role in the organization of the language faculty. Claims of efficient computation require an independently motivated metric of complexity. To this end, we conjecture that the mental representation of morphological knowledge is driven by the time complexity of online processing: productivity is the result of maintaining an optimal balance between lexical and productive rules.

Though direct storage of derived forms has been suggested (Baayen, 2003),[16] the combinatorial explosion of morphologically complex languages (Hankamer, 1989; Niemi et al., 1994; cf. Chan, 2008) necessitates a stage-based architecture of processing that produces morphologically complex forms by rule-like processes (Caramazza, 1997; Levelt et al., 1999). At the minimum, the stem must be retrieved from the lexicon and then combined with appropriate rules/morphemes to generate the derived form. Both processes appear to be geared toward real-time efficiency, where a telling source of evidence comes from frequency effects. One of the earliest and most robust findings in lexical processing is that high frequency words are recognized and produced faster than low frequency words in both visual and auditory tasks (Forster and Chambers, 1973; Balota and Chumbley, 1984). Within the component of morphological computation, it is well established that the processing of exceptions (e.g., irregulars) is strongly correlated with their frequency (see Pinker and Ullman, 2002 for reviews). Such findings have been considered problematic for discrete representations in generative morphology – see Seidenberg and Gonnerman (2000), Hay and Baayen (2006) – but that would be far too hasty; see Yang (2008) for general discussion of probabilistic matters in generative grammar. When understood in terms of modern computer science algorithms, formal models of linguistic competence can be directly translate into a performance model (cf. Miller and Chomsky, 1963; Berwick and Weinberg, 1984); they not only provide accommodation for behavioral results such as frequency effects but also lead to an evaluation measure for productivity.

Generative theories traditionally hold that the organization of morphology is governed by the Elsewhere Condition (Kiparsky, 1973; Halle, 1990), which requires the application of the most specific rule/form when multiple candidates are possible. This provides a way for representing exceptions together with rule-following items. Algorithmically, the Elsewhere Condition may be implemented as a serial search procedure[17]:

[15] For the uninitiated, this is American Baseball talk, referring to the percentage of a batter managing a hit. Batters with sufficiently high batter averages get rich and famous, and the rest languish in the Minors.

[16] As noted by Pinker (1999), there has been a lack of reports of storage effects in auditory presentations of regularly inflected items; it is thus possible that Baayen's results may be an artifact of orthographic familiarity.

[17] This is again a return to an earlier approach in lexical processing, the serial search model of Forster (1976, 1992). The advantage of this model lies in its ready availability for analytic methods, and its empirical coverage is at least as good as other approaches (Murray and Forster, 2004).

(10) IF $w = w_1$ THEN . . .
IF $w = w_2$ THEN . . .
. . .
IF $w = w_M$ THEN . . .
apply R

If taken literally, the Elsewhere Condition treats exceptions as a list of if-then statements that must be evaluated and rejected before reaching the rule-following words. For example, suppose $W_{1...m}$ are m irregular verbs of English: to inflect a regular verb, the language user must determine that it is not one of the irregulars, which would have triggered more specific rules/forms, before the application of the ''add -d'' rule. Now the gradient frequency effects in morphological process can directly captured in the search algorithm. For instance, if the exception clauses (1–m) in (10) are ordered with respect to their token frequencies of use, then the time required to access a specific entry will directly correlate with its position on the list: more frequent entries will be placed higher on the list and will thus be accessed faster. And the construction of a frequency ranked list can be achieved by online algorithms that carry minimal computational costs and thus psychological plausibility. For instance, a simple algorithm Move-Front that moves the most recently accessed clause to the beginning of the list can be shown to be no worse than the optimal list by a small constant factor (Rivest, 1976).

The key feature of the model in (10) is that productivity doesn't come for free: a productive rule may induce considerable time complexity in on-line processing. Specifically, the computation of the regulars will have to ''wait'' until all irregular clauses are evaluated and rejected. Thus, if a rule has too many irregulars (as measured by type frequency), the overall complexity of morphological computation may be slowed down. An immediate prediction of such a model goes 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. The model in (10) entails that w_e will be computed faster than w_r *if* the following conditions are met:

(11) a. the lexical/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.[18]

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 and Mindt, 1995). To the best of my knowledge, 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, and there are four other classes with varying degrees of productivity (Marcus et al., 1995).[19] 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., Sonnestuhl and Huth, 2002). Lexical decision tasks show that when -er and -s suffixing stems are matched in frequency, the -s words show considerably slower reaction time Penke and Krause, 2002; Sonnestuhl and Huth, 2002. The other test 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. These pieces of evidence, along with the treatment of frequency effects among exceptions, provide empirical support for the processing model motivated by the Elsewhere Condition (10).

[18] A frequently used rule will be processed faster than a less frequently used one; this effect can be observed in both morphological processing (Sereno and Jongman, 1997) and morphological learning (Yang, 2002; Mayol and Yang, 2008).

[19] The matters are more complicated than the regular vs. irregular dichotomy portrayed in the dual route approach; see Wiese (1996), Wunderlich (1999) for more refined descriptions of German noun morphology and Yang (2005) for empirical evidence of why at least some of the ''irregular'' classes must be productive themselves.

We now turn to the formal properties of the processing model. Consider 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 (10). Let $T(N, M)$ be the expected time of processing 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 processing for a list thus organized. We conjecture:

(12) 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 $(N - M)$ 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. And the expected time can be expressed as the rule search time weighted by word frequencies. Under the assumption that word frequencies follow the Zipfian distribution, it is possible to show that

(13) **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. Here I will review a simple application of the theorem (13) to provide an explanation for the phenomenon of overregularization.

Following Marcus et al. (1992), let us take the first instance of over-regularization as the onset of the productive use of the ‘‘add -d’’ rule. (Prior to this point, regular past tense forms produced by children are presumably lexicalized, just like the irregular verbs.) (13) implies that at this particular juncture, the learner must know a great deal more regular verbs than irregulars. It is difficult to obtain precise measures of the child's vocabulary, let alone at some specific point: the following is what we have managed for Adam, the boy in Brown's (1973) study. The first instance of Adam's overregularization is at 2;11: ‘‘What dat feeled like?’’: we thus examined his speech from the first transcript recorded at 2;3 to this point. We counted any inflected form of a verb in Adam's production to be part of his vocabulary. There are 211 verbs altogether, among which 143 are regular; the percentage of regular verbs is 68%. According to (13), the predicted percentage ought be

(14)

$$[211 - \frac{211}{\ln(211)}]/211 = 0.81$$

Obviously there is a difference between the expected value and the test but that is merely a sampling effect—regular verbs are much lower in token frequency than irregular vebrs, and are surely more likely to be undersampled. Most importantly, the number of regulars overwhelms the irregulars when the ‘‘add -d’’ rule becomes productive, which is consistent with the general direction of the theoretical model.

A brief note about the terminology – and the scope of our investigation – before we proceed. The representation in (10) might give the impression that R only concerns ‘‘Elsewhere’’ rules, i.e., the default rule that could apply across the board. That is not the case. We are broadly interested in the productivity of any rule whose structural description may include a subset of exceptions. Indeed, rules may be ‘‘nested’’ like Venn diagrams, with the outer rules having broader structural descriptions than inner ones, and each rule may contain listed exceptions.

Nested rules are certainly attested in languages. For instance, it could be said that to the extent that English derivational morphology has a default nominalization process, it is the suffixation of *-ness*; this corresponds to *I* in Fig. 1. However, there are subclasses that are perfectly productive: adjectivals ending /əbəl/ add the suffix *-ity*, which corresponds to *N* above. Rules in nest structures, of course, may or may not have exceptions: In the present discussion, the inner *-ity* rule under /əbəl/ does not seem to have exceptions but the outer *-ness* rule has plenty (e.g., *grow-growth, govern-government*). The German noun plural system appears to be organized in a similar way; see

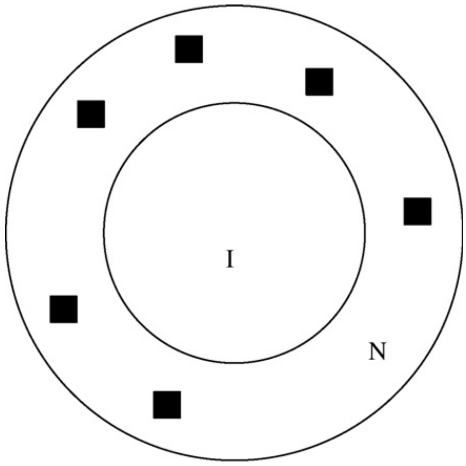

Fig. 1. Rules as nested structures.

footnote 19. We assume application of rules will be from inside out, matching the most specific patterns first as dictated by the Elsewhere Condition (10). For instance, nominalization of an adjectival *A* in English goes as follows. The language user first check if *A* fits the structural description of *I*: if it does, then *I* applies. (If *I* has exceptions, the user will first check if *A* is one of them). If *A* does not match *I*, it will trickle outward to rule *N*. Again, the language user first checks if *A* is one of the exceptions of *N* first, i.e., the black boxes in Fig. 1. If so, those special forms apply; otherwise *I* will step in.

To derive nested rules, and their correct productivity status, is a complicated matter. We assume that the learner can inductively derive linguistic descriptions that are appropriate candidates for productivity calculation; a broad range of learning models are applicable here; see footnote (14). Once these *potentially* productive rules are in place, the child learner will be in a continuous evaluation process that measures and updates their productivities. For instance, the rule ‘‘i→a/__ ŋ’’ we used earlier for expository purposes might indeed have been productive for a young child but quickly demoted as exceptions to it are accumulated. Likewise, the ‘‘add -d’’ rule initially must be unproductive, for the early past tense vocabulary is likely to consist mostly of irregular verbs, which are far higher in frequency than regulars; its promotion to productive status will have to wait for more regular verbs to register, as we have shown in the case of Adam.

Interestingly, there is one child (Abe; Kuczaj, 1976) in the past tense acquisition who did not show the so-called U-shape learning in his longitudinal data: he started overregularization in his first recording session, at 2;4, which is at least six months before Adam (and Sarah; Marcus et al., 1992). The only possible account for this under the productivity model is that Abe must build up the regular portion of his verb vocabulary much faster. Unfortunately, since Abe started overregularization in the very first transcript, it is impossible to estimate his vocabulary composition as we did for Adam. Yet several strands of evidence together confirm the prediction. Abe turned out to be a fantastic word learner, with a Peabody verbal IQ of 140, and he was using more regular verbs than irregulars already at 2;4, where his peers were using considerably more irregulars than regulars at the same age. Moreover, Abe learned more regulars than irregulars in every subsequent month (Maratsos, 2000; Appendix). For his peers, the monthly learning rate for regulars only caught up with that of irregulars at least six months later. In other words, Abe simply reached the productivity threshold a good deal faster than his peers. Such individual level explanations can only be achieved when the learning mechanisms are articulated with sufficient clarity and the learning data are taken seriously as a component in the explanation of child language.

3.2. Evaluating datives

Does the productivity model, which is grounded in morphological learning and processing, generalize to other arenas of linguistic analysis? Here we provide quantitative analysis for the acquisition of syntactic rules.

While the tension between storage and computation has been the focus of morphological processing since the very beginning (Taft and Forster, 1976), there is comparatively less work in this direction for syntactic processing. Fortunately, the important study of Swinney and Cutler (1979) provides experimental results which lend support to the productivity model developed here. Swinney ans Cutler conduct a series of experiments investigating the real time processing complexity of idiomatic expressions such as ''he kicked the bucket'' compared with fully compositional ones such as ''he lifted the bucket''. By controlling for various frequency factors, they find that idioms across a broad range of productivity (Fraser, 1970) are processed significantly faster than compositional expressions. Interpreted algorithmically, the output of parsing (i.e., tree-like structures) is shipped off to the semantic interpretation system, which appears to first perform a lexical lookup for stored structures with idiomatic meanings. If that fails, compositional interpretation would ensue, resulting in processing delay. Note that this process would mirror that of morphological processing discussed earlier, suggesting that the Elsewhere Condition may be a general principle running through many domains of linguistic representation and processing.

With this background in mind, let us turn to the acquisition of the double objective construction and the prepositional *to*-dative construction. Earlier analyses have postulated a derivational relationship between these constructions (Baker, 1988; Larson, 1988), though recent years have seen their treatments as two distinct constructions (Pesetsky, 1995; Harley, 2002 among others). In addition, there is now broad agreement on the semantic conditioning of these constructions (Goldberg, 1995; Hale and Keyser, 2002; Beck and Johnson, 2004, etc.) Adopting the formulations of Pinker (1989), Krifka (1999), the double object dative involves caused possession of the theme by the goal, understood in the broad, including metaphorical, sense (15a) whereas the *to* dative must entail caused motion of the theme along a path to goal (15b):

(15) a. Double object: NP CAUSES NP_{theme}
b. *To* dative: NP CAUSES NP_{theme} TO GO TO NP_{goal}.

But it is important to note that these are no more than necessary conditions on the availability of dative constructions, as language specific constraints further restrict productivity.[20] Korean, for instance, has extremely limited productivity in these constructions (Jung and Miyagawa, 2004). Specifically, Korean exhibits alternations between Dative–Accusative objects and Accusative–Accusative objects, with the former patterning with prepositional datives and the latter with double object datives. The verbs *cwu* (''give''), *kaluch* (''teach'') and *cipwul* (''pay'') can appear in the double object construction but verbs such as *ponay* (''send'') do not. The child learner of Korean, then, will need to learn that verbs do not productively participate in this construction and memorize the handful of exceptions that do. Similar observations can be made about Yaqui, an Uto-Aztecan language spoken in Sonora, Mexico and Arizona. Jelinek and Carnie (2003) show that although most Yaqui ditransitive verbs have an accusative/dative argument array, there is a small closed class of verbs that allow their internal arguments to be marked with two accusative cases, which is equivalent to the English double object construction. Finally, there may be additional constraints within a language that further affect syntactic productivity. For instance, even though the double objective construction is by and large productive in English, it does not extend to Latinate verbs (Green, 1974; Oehele, 1976):

(16) a. John told/reported the news to Bill.
b. John told/*reported Bill the news.

The learning of datives, then, seems to involve the following process. UG provides the candidate verb sets such as (15) with specific semantic properties, but the learner still has to learn whether and which of these verbs can participate in constructions such as in (15a) and (15b). In English, the learner must learn that these constructions are productive—and we find recent extensions to novel verbs which fit the semantic criteria, as in ''I cc'ed everyone the announcement'' and ''I texted the apology to John''. The acquisition of datives is thus similar to that of past tense. In the case of past tense, the general rule has counter examples (i.e., irregulars), and the task is to determine when a sufficient amount of positive examples have been accumulated to counterbalance the exceptions. In the case of English dative acquisition, the constructions in question do not have counterexamples, save restrictions such as (16), but the learner also needs to see a sufficient amount of attested examples to justify the scope of these constructions. In both cases, calibration of linguistic productivity is required.

[20] Another necessary condition may have to do with availability of a certain prepositional element (P_{have}) that encoding possession, as suggested in a decompositional approach to double object constructions (Harley, 2002). In fact, one can frame Harley's position as a problem of morphological productivity; that is, whether the head P_{have} can combine with the relevant stems productively.

The empirical evidence from the acquisition of datives supports this view. Gropen et al. (1989) have noted that English dative acquisition is initially conservative with the child closely following the attested constructions in adult speech. The emergence of productivity is indicated by extension of usage to novel items (c.f., Conwell and Demuth, 2007), including errors that mirror those of overregularization (e.g., Bowerman, 1978) even though the rate of errors is very low (Gropen et al., 1989:209, 217 and the references cited therein)[21]:

(17) a. Adam: Ursula, fix me a tiger.
b. Adam: You finished me a lot of rings.
c. Christy: Don't say me that or you'll make me cry.
d. Ross: Jay said me no.
e. Mark: Ross is gonna break into the TV and is gonna spend us money. (I.e. cause us to spend money for repair.)
f. Damon: Mattia demonstrated me that yesterday.

Now we are in the position to offer a quantitative account of how – or precisely, when – these constructions are learned. As documented in Snyder and Stromswold (1997), the emergence of the double objective construction is significantly ahead of that of the *to*-dative construction. The application of the productivity model (13) is quite straightforward. We examined a random sample of 86,442 child directed utterances from the CHILDES database (MacWhinney, 1995) that was used for a word segmentation project (Yang, 2004). We only considered verb stems that appeared at least 5 times in the input data on the ground that very low frequency verbs may not be firmly placed in a young learner's vocabulary (see Gropen et al., 1989). From these stems, we generated two sets that could, and indeed do, participate in the double object and to-dative constructions. Then, from these two sets, we counted the number of verbs actually attested at least once in the respective constructions.[22] These verbs are exhaustively listed below, where those in bold are attested for the relevant construction:

(18) Double object construction (15a):
get, **tell**, **make**, **give**, **write**, **bring**, **tell**, **read**, sing, **find**, hit, **draw**, **show**, **buy**, throw, **send**, **teach**, **feed**, cook, **buy**, sell, **pay**, **bake**, **serve**, **teach**

(19) *To*-dative construction (15b):
tell, **take**, **give**, **said**, write, **bring**, **read**, sing, hit, play, **throw**, pull, push, **show**, shoot, blow, drive, **send**, carry, teach, feed, fly, move, **pay**, serve.

Out of the 25 candidates for the double object construction, 19 are actually used in the input. By contrast, out of the 25 candidates for the to-dative construction, only 10 are attested in the input. Thus, the batting average for (18) is far higher than that for (19)—in fact, if we apply the theorem in (13), we see that the double object construction has already crossed the threshold of productivity but the *to*-dative construction is nowhere close. Naturally, we expect that as the volume of input data is increased, more and more verbs in (19) will participate in the to-dative construction and the threshold of productivity will eventually be met. But the point here is that the tipping point for the double object construction is reached sooner. This pattern of acquisition is ultimately caused by the statistical distribution of usage in the input; it takes a formal model of productivity and learning to map the quantitative measures of the data to syntactic regularities of the grammar.

4. Conclusion

We conclude with some brief notes on the connection from language acquisition to the theory of grammar.

The core vs. periphery distinction, like parameters, seems to have fallen out of fashion these days. A current trend in linguistic theorizing aims to dispense with the core parametric system and replace it with a plethora of rules and constructions (Culicover, 1999; Newmeyer, 2004; Culicover and Jackendoff, 2005) which are presumably inductively

[21] It remains to be seen how the learner acquires additional restrictions such as (16); see Gropen et al. (1989), Pinker (1989) for discussion.
[22] A verb is considered to have participated in a construction if at least one of its inflectional forms is used in that construction.

learned from the data under the guidance of certain principles.[23] One reason for abandoning the core, I suspect, lies in the lack of an articulated theory that draws boundaries from the periphery. Since the periphery is necessary anyway – no one doubts that idioms and irregular verbs must be learned – what's left for the core? Assuming that something like the productivity model presented here is correct, one might ask, and I have indeed been asked, whether that sort of productivity analysis plus some general inductive learning model can carry the learner all the way.

I don't think it is wise to abandon the core, even if we put aside the utility of the parameter system for comparative linguistic descriptions and focus solely on language acquisition instead. From a formal learnability point of view, a finite space of parameters or constraints is still our best bet on the logical problem of language acquisition (Nowak et al., 2002) but that's for another day. It is useful to recall the distinction between core and periphery originally drawn by Chomsky (1981):

> Marked structures have to be learned on the basis of slender evidence too, so there should be further structure to the system outside of core grammar. We might expect that the structure of these further systems relates to the theory of core grammar by such devices as relaxing certain conditions of core grammar, processes of analogy in some sense to be made precise, and so on, though there will presumably be independent structure as well. (p. 8)
> . . .
> How do we delimit the domain of core grammar as distinct from marked periphery? In principle, one would hope that evidence from language acquisition would be useful with regard to determining the nature of the boundary or the propriety of the distinction in the first place, since it is predicted that the systems develop in quite different ways. (p. 9)

Of course, even if the core vs. periphery boundary is difficult to draw, it does not mean that no such boundary exits. Regardless the merit of the specific learning models presented here, the facts from child language do speak to the two distinct components of the grammar that Chomsky alludes to. One type is documented in the quantitative studies of section 2. The child spontaneously accesses grammatical options for which she has no external evidence, and the elimination of these options is sensitive to **token** frequencies of specific linguistic data that are quite far removed from surface level patterns (e.g., expletives for obligatory subjects). The most straightforward interpretation is an internal one: such non-target grammatical options are part of the innate endowment that is invariant across languages and speakers. Quite a distinct kind can be found in section 3, where the substantive form of linguistic knowledge is derived from the external, language particular, environment, which corresponds to the periphery. Acquisition of the periphery is characterized by the conservativeness of the learner at the initial stage: before productive use of the "add -d" rule and dative constructions, the child only keeps to attested forms in the input and does not generalize (recall especially the asymmetry in overregularization vs. overirregularization in section 3.1). Moreover, the acquisition of the periphery appears to be a process sensitive to the **type** frequency of the relevant linguistic experience: hearing "walked" a thousand times will not give the learner the rule of "add -d" but hearing a sufficiently diverse range of expressions (*walk-walked, kill-killed*, *fold-folded*, etc.) will do, even if there are some interfering irregulars that are heard very frequently. And the periphery isn't a land of the lawless either. Successful acquisition (of anything) will not be possible with structural principles that govern **all** components of the grammars. The interdeterminacy of induction is a matter of logic. When deriving a language specific generalization, the child must be constrained by both linguistic and possible non-linguistic principles of learning: no search for syntactic generalizations over every other word, for example, wherever such generalizations may lie in the linguistic system. The force of the core must radiate through the periphery.

Which leads us to reiterate our position that not asking UG to do too much doesn't mean asking UG to do too little. The mechanism of learning and the representation of linguistic knowledge are in principle quite different matters, but they must fit well together to yield useful results. An analogy can be made to the separation of data structure and algorithms in computer science: an algorithm (e.g., binary search) operates on any tree but would only turn out good performance when the tree is structured in specific ways. The variational model of learning quite possibly derives from evolutionarily ancient learning and decision making mechanisms in other domains and species (Bush and Mosteller, 1951; Atkinson et al., 1965) and similar models have seen applications in other areas of language acquisition (Labov, 1994). But it requires particular partitions of the grammar space – rather than making a left or right turn for the rodent

[23] Which may or may not be specific to language, and may or may not be innate; here opinions differ greatly.

in a maze (Gallistel, 1990) – to provide accurate descriptions of children's syntactic development. Likewise, the principle of efficient computation that motivates the productivity model may be a general constraint on cognitive and neural systems, yet its execution is completely derived from the linguistic principle of the Elsewhere Condition, or whatever independent principle from which the Elsewhere Condition derives. A large range of language variation, then, falls under the learner's ability to derive appropriate generalizations from the data; not having to build everything about language in the genetic endowment may produce a concrete program for the study of language in an evolutionary setting (Hauser et al., 2002).

Acknowledgements

I thank the audiences at the Formal Approach to Variation in Syntax at the University of York, Utrecht University, University of Massachusetts at Amherst, and Princeton University where portions of this paper were presented. Two anonymous reviewers deserve special mention for their very helpful criticisms and suggestions. Thanks are also due to Bernadette Plunkett and Bill Haddican for their patience and guidance.

References

Aronoff, M., 1976. Word formation in generative grammar. MIT Press, Cambridge, MA.

Atkinson, R., Bowever, G., Crothers, E., 1965. An Introduction to Mathematical Learning Theory. Wiley, New York, NY.

Baayen, H., 2003. Probabilistic approaches to morphology. In: Bod, R., Hay, J., Jannedy, S. (Eds.), Probabilistic Linguistics. MIT Press, Cambridge, MA, pp. 229–287.

Baayen, H., Lieber, R., 1991. Productivity and English derivation: a corpus-based study. Linguistics 29, 801–843.

Baker, M., 1988. Incorporation. University of Chicago Press, Chicago, IL.

Baker, M., 2002. The Atoms of Language. Basic Books, New York.

Balota, D., Chumbley, J., 1984. Are lexical decisions a good measure of lexical access? The role of frequency in the neglected decision stage. Journal of Experimental Psychology: Human Perception and Performance 10, 340–357.

Bates, E., 1976. Language and Context: the Acquisition of Pragmatics. Academic Press, New York, NY.

Beck, S., Johnson, K., 2004. Double objects again. Linguistic Inquiry 1, 97–124.

Berko, J., 1958. The child's learning of English morphology. Word 14, 150–177.

Berwick, R., Niyogi, P., 1996. Learning from triggers. Linguistic Inquiry 27, 605–622.

Berwick, R., Weinberg, A., 1984. The Grammatical Basis of Linguistic Performance. MIT Press, Cambridge, MA.

Bloom, P., 1990. Subjectless sentences in child languages. Linguistic Inquiry 21, 491–504.

Bloom, P., 1993. Grammatical continuity in language development: the case of subjectless sentences. Linguistic Inquiry 24, 721–734.

Bowerman, M., 1978. Systematizing semantic knowledge: Changes over time in the child's organization of word meaning. Child Development 49 (4), 977–987.

Bromberger, S., Halle, M., 1989. Why phonology is different. Linguistic Inquiry 20 (1), 51–70.

Brown, R., 1973. A First Language. Harvard University Press, Cambridge, MA.

Bush, R., Mosteller, F., 1951. A mathematical model for simple learning. Psychological Review 68, 313–323.

Caramazza, A., 1997. How many levels of processing are there in lexical access? Cognitive Neuropsychology 14 (1), 177–208.

Chan, E., 2008. Distributional and Structural Basis on Morphological Learning. Ph.D. Dissertation. Department of Computer Science, University of Pennsylvania, Philadelphia, PA.

Chomsky, N., 1955. The Logical Structure of Linguistic Theory. Manuscript. Harvard University and Massachusetts Institute of Technology.

Chomsky, N., 1965. Aspects of the Theory Of Syntax. MIT Press, Cambridge, MA.

Chomsky, N., 1977. On Wh-movement. In: Culicover, P., Wasow, T., Akmajian, A. (Eds.), Formal Syntax. Academic Press, New York, pp. 71–132.

Chomsky, N., 1981. Lectures on Government and Binding. Foris, Dordrecht.

Chomsky, N., 2005. Three factors in language design. Linguistic Inquiry 36 (1), 1–22.

Chomsky, N., Halle, M., 1968. The Sound Patterns of English. MIT Press, Cambridge, MA.

Clahsen, H., 1986. Verbal inflections in German child language: acquisition of agreement markings and the functions they encode. Linguistics 24, 79–121.

Clahsen, H., Hadler, M., Weyerts, H., 2004. Speeded production of inflected words in children and adults. Journal of Child Language 31, 683–712.

Conwell, E., Demuth, K., 2007. Early syntactic productivity: evidence from dative shift. Cognition 2, 163–179.

Crain, S., Pietroski, P., 2002. Why language acquisition is a snap. Linguistic Review 19, 163–183.

Culicover, P., 1999. Syntactic Nuts. Oxford University Press, Oxford.

Culicover, P., Jackendoff, R., 2005. Simpler Syntax. Oxford University Press, Oxford.

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

Forster, K., 1992. Memory-addressing mechanisms and lexical access. In: Frost, R., Katz, L. (Eds.), Orthography, Phonology, Morphology, and Meaning. Elsevier, Amsterdam, pp. 413–434.

Forster, K., Chambers, S., 1973. Lexical access and naming time. Journal of Verbal Learning and Verbal Behavior 12, 627–635.

Fraser, B., 1970. Idioms in transformational grammar. Foundations of Language 6, 22–43.

Gallistel, C.R., 1990. The Organization of Learning. MIT Press, Cambridge, MA.

Gerken, L.A., 1991. The metrical basis for children's subjectless sentences. Journal of Memory and Language 30, 431–451.

Gibson, E., Wexler, K., 1994. Triggers. Linguistic Inquiry 25, 355–407.

Goldberg, A., 1995. Constructions. University of Chicago Press, Chicago, IL.

Golinkoff, R., Hirsh-Pasek, K., Cauley, K., Gordon, L., 1987. The eyes have it: Lexical and syntactic comprehension in a new paradigm. Journal of Child Language 14, 23–46.

Grabowski, E., Mindt, D., 1995. A corpus-based learning of irregular verbs in English. Computers in English Linguistics 19, 5–22.

Green, G., 1974. Syntactic and Semantic Regularity. Indiana University Linguistics Club, Bloomington, IN.

Gropen, J., Pinker, S., Hollander, M., Goldberg, R., Wilson, R., 1989. The learnability and acquisition of the dative alternation in English. Language 65 (2), 203–257.

Guasti, M., 2002. Language Acquisition: The Growth of Grammar. MIT Press, Cambridge, MA.

Haegeman, L., 1995. Root infinitives, tense, and truncated structures. Language Acquisition 4, 205–255.

Hale, K., Keyser, S.J., 2002. Prolegomenon to a Theory of Argument Structure. MIT Press, Cambridge, MA.

Halle, M., 1990. An approach to morphology. Proceedings of the Northeast Linguistic Society 20, 150–184.

Halle, M., Mohanan, K.-P., 1985. Segmental phonology of modern English. Linguistic Inquiry 16, 57–116.

Hankamer, J., 1989. Lexical representation and processes. In: Marslen-Wilson, W. (Ed.), Morphological Parsing and the Lexicon. MIT Press, Cambridge, MA.

Harley, H., 2002. Possession and the double object construction. Language Variation Yearbook 2, 29–68.

Hauser, M., Chomsky, N., Fitch, T., 2002. The faculty of language: what is it, who has it, and how did it evolve? Science 298 (5598), 1569–1579.

Hay, J., 2003. Causes and Consequences of Word Structure. Routlege, New York.

Hay, J., Baayen, H., 2006. Shifting paradigms: gradient structure in morphology. Trends in Cognitive Science 9 (7), 342–348.

Hyams, N., 1986. Language Acquisition and the Theory of Parameters. Reidel, Dordrecht.

Hyams, N., Wexler, K., 1993. On the grammatical basis of null subjects in child language. Linguistic Inquiry 24, 421–459.

Jakobson, R., 1941/1968. Child Language, Aphasia and Phonological Universals. Mouton, The Hague.

Jelinek, E., Carnie, A., 2003. Argument hierarchies and the mapping principle. In: Carnie, A., Harley, H., Willie, M. (Eds.), Formal Approaches to Function. John Benjamins, Philadelphia, pp. 265–296.

Jung, Y.-J., Miyagawa, S., 2004. Decomposing ditransitive verbs. In: Proceedings of SICGG. pp. 101–120.

Kiparsky, P., 1973. "Elsewhere" in phonology. In: Anderson, S., Kiparsky, P. (Eds.), A Festschrift For Morris Halle. Holt, Rinehart, and Winston, New York, pp. 93–106.

Krifka, M., 1999. Manner in Dative Alternations. In Proceedings of WCCFL 18. Cascadilla Press. Somerville, MA. pp. 260–271.

Kuczaj, S., 1976. "-ing", "-s", and "-ed": a study of the acquisition of certain verb inflections. Unpublished Doctoral Dissertation. University of Minnesota, Department of Child Psychology.

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.

Labov, W., 1994. Principles of Language Change: Internal Factors. Blackwell, Basil.

Larson, R., 1988. On the double object construction. Linguistic Inquiry 19 (3), 335–391.

Legate, J.A., Yang, C., 2007. Morphosyntactic learning and the development of tense. Language Acquisition 14 (3), 315–344.

Levelt, W.J.M., Roelofs, A., Meyer, A., 1999. A theory of lexical access in speech production. Behavioral and Brain Sciences 1, 1–75.

Lewontin, R., 1983. The organism as the subject and object of evolution. Scientia 118, 65–95.

Lightfoot, D., 1999. The Development of Language. Blackwell, Basil.

MacWhinney, B., 1995. The CHILDES Project: Tools for Analyzing talk. Lawrence Erlbaum, Mahwah, NJ.

Maratsos, M., 2000. More overregularizations after all: new data and discussion on Marcus, Pinker, Ullman, Hollander, Rosen, & Xu. Journal of Child Language 27, 183–212.

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.

Mayol, L., Yang, C., 2008. Regularities in Spanish Irregular Morphology. Manuscript. Department of Linguistics, University of Pennsylvania, Philadelphia, PA.

Miller, G., Chomsky, N., 1963. Introduction to the Formal Analysis of Natural Languages. In: Luce, D., Bush, R., Galanter, E. (Eds.), Handbook of Mathematical Psychology II. Wiley and Sons, New York, pp. 268–321.

Mitchell, T., 1982. Generalization as search. Artificial Intelligence 18, 203–226.

Murray, W., Forster, K., 2004. Serial mechanisms in lexical access: the rank hypothesis. Psychological Review 111, 721–756.

Newmeyer, F., 2004. Against a parameter-setting approach to language variation. Language Variation Yearbook 4, 181–234.

Niemi, J., Laine, M., Tuominen, J., 1994. Cognitive morphology in Finnish: foundations of a new model. Language and Cognitive Processes 9, 423–446.

Nowak, M., Komarova, N., Niyogi, P., 2002. Computational and evolutionary aspects of language. Nature 417, 611–617.

Oehele, R., 1976. The Grammatical Status of the English Dative Alternation. MIT Dissertation. Cambridge, MA.

Penke, M., Krause, M., 2002. German noun plurals: a challenge to the dual-mechanism Model. Brain and Language 81 (1–3), 303–311.

Pesetsky, D., 1995. Zero Syntax: Experiencer and Cascade. MIT Press, Cambridge, MA.

Phillips, C., 1995. Syntax at age 2: cross-linguistic differences. In: MIT Working Papers in Linguistics 26. MITWPL, Cambridge, MA, pp. 325–382.

Pierce, A., 1992. Language Acquisition and Syntactic Theory. Kluwer, Boston, MA.
Pinker, S., 1984. Language Learnability and Language Development. Harvard University Press, Cambridge, MA.
Pinker, S., 1989. Language Learnability and Cognition. MIT Press, Cambridge, MA.
Pinker, S., 1999. Words and Rules: The Ingredients of Language. Basic Books, New York, NY.
Pinker, S., Ullman, M., 2002. The past and future of past tense. Trends in Cognitive Sciences 11 (1), 456–463.
Poeppel, D., Wexler, K., 1993. The full competence hypothesis. Language 69, 1–33.
Rappaport Hovav, M., Levin, B., 2008. The English dative alternation: The case for verb sensitivity. Journal of Linguistics 44, 129–167.
Rivest, R., 1976. On self-organizing sequential search heuristics. Communications of the ACM 2, 63–67.
Rizzi, L., 2004. Grammatically Based Target Inconsistencis in Child Language. Manuscript. University of Siena.
Roeper, T., 2000. Universal bilingualism. Bilingualism: Language and Cognition 2, 169–186.
Saffran, J.R., Aslin, R.N., Newport, E.L., 1996. Statistical learning by 8-month old infants. Science 274, 1926–1928.
Sakas, W., Fodor, J.D., 2001. The structural triggers learner. In: Bertolo, S. (Ed.), Language Acquisition and Learnability. Cambridge University Press, Cambridge, pp. 172–233.
Schütze, C., 2005a. Thinking about what we are asking speakers to do. In: Kepser, S., Reis, M. (Eds.), Linguistic Evidence: Theoretical, Empirical, and Computational Perspectives. Mouton de Gruyter, Berlin, pp. 457–485.
Schütze, C., 2005b. Thinking About What We Are Asking Speakers To Do. Manuscript. UCLA.
Seidenberg, M., Gonnerman, L., 2000. Explaining derivational morphology as the convergence of codes. Trends in Cognitive Sciences 4, 353–361.
Sereno, J., Jongman, A., 1997. Processing of English inflectional morphology. Memory and Cognition 25, 425–437.
Snyder, W., 2007. Child Language: The View from Parameters. Oxford University Press, Oxford.
Snyder, W., Stromswold, K., 1997. The structure and acquisition of english dative constructions. Linguistic Inquiry 28 (2), 281–317.
Sonnestuhl, I., Huth, A., 2002. Processing and representation of german -n plurals: a dual-mechanism approach. Brain and Language 81 (1–3), 276–290.
Straus, K., 2008. Validations of a probabilistic model of language learning. Ph.D. Dissertation. Department of Mathematics, Northeastern University, Boston, MA.
Stromwold, K., Zimmerman, K., 1999. Acquisition of *Nein* and *Nicht* and the VP-internal subject stage in German. Language Acquisition 8, 101–127.
Sussman, G., Yip, K., 1997. Sparse representations for fast, one-shot learning. In: Paper Presented at the National Conference on Artificial Intelligence, Orlando, FL.
Swinney, D., Cutler, A., 1979. The access and processing of idiomatic expressions. Journal of Verbal Learning and Verbal Behavior 18 (5), 523–534.
Taft, M., Forster, K., 1976. Lexical storage and retrieval of polymorphemic and polysyllabic words. Journal of Verbal Learning and Verbal Behavior 15 (6), 607–620.
Tomasello, M., 2003. Constructing a Language: A Usage-based Approach to Language. Harvard University Press, Cambridge, MA.
Valian, V., 1991. Syntactic subjects in the early speech of American and Italian Children. Cognition 40, 21–82.
Wang, Q., Lillo-Martin, D., Best, C., Levitt, A., 1992. Null subject vs. null object: some evidence from the acquisition of Chinese and English. Language Acquisition 2, 221–254.
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.
Wexler, K., 1998. Very early parameter setting and the unique checking constraint: a new explanation of the optional infinitive stage. Lingua 106, 23–79.
Wiese, R., 1996. The phonology of German. Cambridge University Press, Cambridge.
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. Oxford University Press, New York.
Yang, C., 2004. Universal grammar, statistics, or both. Trends in Cognitive Sciences 8 (10), 451–456.
Yang, C., 2005. On productivity. Language Variation Yearbook 5, 333–370.
Yang, C., 2006. The Infinite Gift. Scribner's, New York.
Yang, C., 2008. The great number crunch. Journal of Linguistics 44, 205–228.