A formalist perspective on language acquisition

Charles Yang

Language acquisition is a computational process by which linguistic experience is integrated into the learner's initial stage of knowledge. To understand language acquisition thus requires precise statements about these components and their interplay, stepping beyond the philosophical and methodological disputes such as the generative vs. usage-based approaches. I review several mathematical models that have guided the study of child language acquisition: How learners integrate experience with their prior knowledge of linguistic structures, How researchers assess the progress of language acquisition with rigor and clarity, and How children form the rules of language even in the face of exceptions. I also suggest that these models are applicable to second language acquisition (L2), yielding potentially important insights on the continuities and differences between child and adult language.

1. A formal introduction

Don Knuth, perhaps the most renowned living computer scientist, infamously took *Syntactic Structures* on his honeymoon. For many language scientists, and for many more outside of linguistics, generative grammar initiated an exciting way of studying a quintessential aspect of human life. It was a refreshing change from the routine in traditional social and behavioral sciences: take measurements, fit a curve, repeat. While the study of language has an ancient tradition, it is safe to say that it was the commitment to a causal and mechanical account of language helped establish the field of cognitive science; the mind may be studied with deductive methods as in the natural sciences.

This article provides a personal perspective on the formalist approach to language acquisition. I use the term "formalist" in the methodological sense: it reflects a commitment to the mechanistic study of language, rather than the conventionalized standin for "generative" and/or "nativist". In my opinion,

https://doi.org/10.1075/lab.18014.yan
Linguistic Approaches to Bilingualism 8:6 (2018), **665–706**.
ISSN 1879–9264 | E-ISSN 1879–9272

methodological rigor should override the researcher's preference for the mode of explanation. (Indeed, as someone trained in the generative tradition which emphasizes domain-specific knowledge, I have been invigorated by usage-based research which places greater emphasis on the role of data-driven learning). More specifically, a formal theory of language acquisition ought to be more than a description, that children know A at age X but B at age $X + Y$ (e.g., the transition from an item-based to a rule-based grammar), or that some variable P (e.g., input frequency) is correlated with some other variable Q (e.g., the correct rate of morphological marking). The theory should also include a mechanistic account of how the $A \rightarrow B$ transition takes place and how P causally affects Q. After all, it was such a mechanistic approach that got everyone excited about linguistics and helped build connections with other fields.

In the spirit of being formal, I have organized this paper around three simple mathematical models. As I will review in Section 2, 3, and 4, they have been developed to address the so-called logical problem of language acquisition (Hornstein & Lightfoot, 1981), that the language learner must go beyond their linguistic experience to form general linguistic rules. In Section 5, I submit, with some trepidation, that these equations may also prove useful for L2 acquisition. This is because learning a second language must also go beyond the data to form generalizations – the logical problem of L2 acquisition (White, 1985; Bley-Vroman, 1989). The equations provide very concrete predictions that can be easily confirmed or disconfirmed. To the extent they are confirmed, we may detect potential continuities between child and adult language acquisition. To the extent they are disconfirmed, we may be a step closer to understanding why children seem better at language learning than adults.

2. Competition and selection

2.1 Background

There was a time when formal methods were integral to the empirical research on child language: I have in mind the influential work of Suppes (1974), Pinker (1979), Wexler and Culicover (1980), Berwick (1985), among others. Formal models make explicit statements about the learner's predisposition for language, the ecological condition of language acquisition (e.g., no negative evidence), and the learning algorithms likely within children's computational capacity (e.g., incremental learning). The commitment to a formal account of language learnability in fact inspired the modern study of machine learning and statistical inference (Solomonoff, 1964; Gold, 1967; Blum & Blum, 1975; Angluin, 1980).

The scene had already changed when I started working on child language in the late 1990s. The field, at least the part of the field I was embedded in, had shifted primarily to developmental issues. A major debate at the time was whether the differences between child and adult language are due to competence or performance gaps (e.g., Pinker, 1984; Borer & Wexler, 1987; Demuth, 1989; Bloom, 1990; Valian, 1991; Wang et al., 1992; Hyams & Wexler, 1993). But both sides of the debate agreed that the child's grammar is already target-like as soon as it can be directly tested (e.g., at the beginning of multi-word combinations). This was a natural position for the performance-based theorists but seemed paradoxical under the competence-based account: if children have exquisite knowledge of their grammar (Wexler, 1998), how come they don't talk right? Conspicuously absent is a learnability account of *how* children's grammar becomes target-like: both sides in effect deny the role of input and experience, because language-specific data has no explanatory power on the acquisition of the grammar.[1]

As someone on the outside looking in, this state of affairs was fascinating but also puzzling. The quantitative work that began to emerge, thanks in no small part to the CHILDES project (MacWhinney, 2000), confirmed that many aspects of child language are indeed adult-like from very early on, as Roger Brown recognized long ago (1973, p. 156). A highlight was the near-perfect correlation between the position and inflection of the verb in the main clause (Pierce, 1992, etc.), which poses significant challenges for adult L2 learners (White, 1990). At the same time, the problem of how children learn their specific grammars was left unaddressed.

On top of this, and perhaps because of it, there was a widely held belief that the commitment to Universal Grammar, shared broadly by both sides of the competence/performance divide, is inherently incompatible with input effects in language acquisition (e.g., Tomasello, 2003, p. 97, Hoff, 2014, p. 106). Distributional learning from data was viewed as evidence against Universal Grammar; see, for instance, "learning rediscovered" (Bates & Elman, 1996) after the discovery of statistical learning for word segmentation (Saffran et al., 1996; see also Yang, 2004). The same period also witnessed the so-called English past tense debate (Pinker & Ullman, 2002; McClelland & Patterson, 2002). Here the disagreement was on the treatment of the regular verbs – whether it is associatively based or whether it is handled by a rule that adds "-ed". For both sides, irregular past tense was formed by associative memory thus sensitive to frequency effects, apparently incompatible with the symbolic treatment of irregulars throughout the history of linguistics (Bloch, 1947; Chomsky & Halle, 1968).

1. Under some accounts, children's output is constrained by performance filters that are themselves subject to language variation (e.g., Gerken, 1994; Demuth, 1996), but these accounts also assert the correctness of child's competence grammar.

2.2 The variational model

It was in this context that I proposed the variational learning model (Yang, 2002). It was an acknowledgment that input effects matter for language acquisition but are completely consistent with the theory of Universal Grammar. It was also a return to the formalist tradition of language research. Rejecting the dominant view that the child language is characterized by a single grammar (e.g., an adult-like grammar, or a grammar in the space of possible grammars as in the Principles and Parameters framework and Optimality Theory), the variational model assumes that the grammars in the child's hypothesis space are associated with probabilities or weights. Learning takes place not by changing one grammar to another (e.g., Wexler & Culicover, 1980; Berwick, 1985; Gibson & Wexler, 1994) but as changes in the probabilistic distribution of the grammars in response to input data. The simplest instantiation of the variational model is the Linear Reward Penalty scheme (Bush & Mosteller, 1951), one of the oldest and best supported models from mathematical psychology.

For the purpose of illustration, consider a learner who has access to two grammars, the target *A* and a competitor *B*, which are currently associated with probabilities p and q. Upon encountering an input item s, the learner selects a grammar with its associated probability. Suppose *A* is chosen:

(1) a. If *A* can analyze s then $p' = p + \gamma q$ and $q' = (1 - \gamma)q$
  b. If *A* cannot analyze s then $p' = (1 - \gamma)p$ and $q' = q + \gamma p$

The chosen grammar has its probability increased if successful and decreased if not: in a zero-sum game, its competitor's probability adjustment is just the opposite. Several remarks about the variational model are in order.

First, the competition scheme in the variational model, inspired by evolutionary models of biological change (Yang, 2006), implies some notion of fitness, and it is the fitness differential of the grammars that drives learning. In the simplest case, the target grammar *A* by definition always succeeds: p will rise whenever *A* is selected. The competitor *B*, by definition, must fail on a certain proportion of the input: when that happens, q will decrease and p will thus increase. But importantly, *B* needn't fail all the time: there may be input items that are ambiguous between the grammars. Thus, the trajectory of learning in the long run is determined by the statistical composition of the input data. A grammar whose competitor is penalized more often will be learned faster.

Second, the grammar-input compatibility, referred to as "analyze" in (1), can be flexibly defined as long as it is precise and independently motivated. The simplest case would be parsability (e.g., whether the grammar is compatible with an input string) but many other considerations are possible. For instance, if the child's parsing system has certain limitations (Trueswell et al., 1999), then even sentences compatible with the target grammar may fail to be analyzed. The fitness values may also socially conditioned: a stigmatized variant would put its competitors at an advantage.[2]

Third, the variational model does not require that the hypotheses in competition are innately available. In fact the formalism is applicable to any finite set of hypotheses, including hypotheses that the learner constructs on the basis of specific language input. For instance, the model has been applied to word learning to represent the probabilistic association between the phonological form of a word and its meaning (Stevens et al., 2016), both of which are clearly learned from the environment. However, the explanatory value of the variational model for specific cases of language acquisition lies in the specific hypotheses under competition, which may provide direct evidence for the nativist position as I discuss later in this section.

Finally, the variational model leaves space for individual variation. The statistical composition of the input may vary such that the target grammar may develop along different schedules for individual learners. It is also possible that some children are just slower at absorbing linguistic input than others; this is operationalized by the learning rate parameter y in (1), which represents the magnitude of probability adjustment as the result of analysis, again a familiar notion from the mathematical psychology of learning. It has been suggested that individual variation in γ is a source for developmental language delay (Legate & Yang, 2007).

The variational learning model was originally applied to the problem of parameter setting: for *A* and *B* in (1), think of the opposite values of a parameter. The model provably converges on the target grammar in the limit (Straus, 2008). In a complex domain of thirteen word-order parameters (Sakas & Fodor, 2012), the variational model has been shown to converge on the target consistently and efficiently (Sakas et al., 2017). More important, the variational model resolves several major challenges associated with traditional approaches to parameter setting. Its probabilistic nature means that the target grammar will only gradually rise to dominance under the cumulative effect of unambiguous data in its favor. Two empirical consequences follow.

First, it is possible to establish the amount of unambiguous evidence for parameter values in child-directed input corpora to correlate with the developmental time course of the parameters. For instance, languages differ in the positioning

2. All the same, it is important to recognize that the fitness value, e.g., the probability with which a grammar fails to analyze the input data, is not something the learner needs to calculate – no more than the mouse needs to tabulate the probabilities of receiving food pellets in conditioning experiments (Bush & Mosteller, 1951).

of the main verb in the matrix clause: for languages like English, the verb follows adverbs (e.g., *John often drinks coffee*) whereas for languages like French, the verb precedes adverbs (e.g., *Jean boit souvent du café).* Only sentences that contain positional signposts such as the adverb *often/souvent* can unambiguously nudge the learner toward their language-specific option (White, 1990); see Yang (2012) for a review of such input effects of parameters across languages. Recent work has devised intervention strategies (Hadley & Walsh, 2014; Hadley et al., 2017) to boost English children's development of the morphosyntax of tense by amplifying the volume of informative data such as third-person singular present tense verbs in the caretaker input (Legate & Yang, 2007; Yang et al., 2015).

Second, under the variational model, children's systematic deviation from the target grammar may be attributed to non-target hypotheses before their eventual demise; see Crain et al. (2016) for a recent review. Naturally, this perspective is only as good as what we take to be the space of linguistically possible hypotheses available to a child. While few would claim that human languages can vary arbitrarily, there is still considerable debate whether such constraints are specific to language or result from the constellation of other cognitive factors. At the same time, the theory of parameters is currently under review even within generative linguistics, especially in light of evolutionary considerations: that properties of human language previously thought to be domain -specific may be ultimately grounded in other cognitive and perceptual systems (Hauser et al., 2002) and the principle of efficient computation (Chomsky, 2005; Berwick & Chomsky, 2016; Yang et al., 2017). In the concluding section of this paper, I will briefly discuss how the Tolerance Principle (Yang, 2016), which is reviewed in Section 4, reduces the explanatory burden traditionally placed on innate linguistic parameters.

In my opinion, a theory of language acquisition need not be overly bound to the latest theory of linguistic structures. Child language can often be fruitfully studied as the level of empirical generalization which, if sufficiently robust, cannot only withstand the changing theoretical perspectives but also actively constrain them. Here I review one specific line of evidence uncovered through the variational model.

2.3 Null subject: The last parameter?

Consider the classic problem of null arguments in child English (Bloom, 1970). English-learning children frequently omit subjects – up to 30% of the time – and they occasionally omit objects as well, quite contrary to the input data they hear. (2) provides some naturally occurring examples from the CHILDES database.

(2) _ want cookies.
Where _ going?
How _ wash it?
Erica took _.
I put _ on.

These missing arguments generally do not impede language understanding. Nevertheless, children do not start using subjects and objects consistently at adult level until around the third birthday. Earlier attempts to equate the null argument stage to parameter missetting to the pro-drop or topic-drop option (Hyams, 1986, 1991) were unsuccessful: during no stage of acquisition does the distribution of English-learning children's argument use resemble that of speakers or learners of pro-drop and topic-drop languages (Valian, 1991; Wang et al., 1992), nor is there any evidence for sudden changes in the frequency of null subjects which would have supported the notion of parameter resetting (Bloom, 1990; Legate & Yang, 2007).

The variational model offers a new perspective on the null argument phenomenon. First, one needs to find distributional evidence in child language for the non-target grammars not yet unlearned. Second, one needs to quantify the disambiguating data in the input that eliminates these non-target grammars. More specifically, when English-learning children probabilistically access the target grammar, they will not generally omit the argument. But when the topic-drop grammar is accessed, argument omission would be possible when the discourse conditions are met. The most telling evidence can be found in a striking distributional property in child English. It is easy to find hundreds of child English examples in the CHILDES corpus of the following type:

(3) a. When _ bring the bag back? When _ rains?
b. Where _ get these? Where _ go? Why _ go slowly?
c. Why _ get scratched by the cat? Why _ laughing at me?
d. How _ fix my eye? How _ do open it?

These questions have target-like fronting of the wh-word but the subject is missing. Notice that these are all *adjunct* questions with *when, where, why,* and *how*. By contrast, omitted subjects in *argument* wh-questions are vanishingly rare:

(4) a. *Who$_t$ _ kissing *t*?
b. *Who$_t$ _ see *t*?
c. *What$_t$ _ want to hit *t*?

That is, when the object (*who* and *what*) is fronted in a wh-question, the subjects are almost never omitted. An exhaustive search of wh-questions produced by Adam, a prolific subject dropper, reveals a near categorical asymmetry (Yang, 2002, p. 120):

(5) a. 95% (114/120) of the wh-questions with an omitted subject are adjunct questions.
b. 97.2% (209/215) of the wh-object questions contain subjects.

The null subject asymmetry in English-learning children's argument and adjunct questions is exactly mirrored in topic-drop languages. Consider the contrast between (6a) and (6b) in Mandarin Chinese. In both cases, suppose the existing discourse topic is the subject "John" but a new topic has been introduced via topicalization (in italic and marked with trace). The omission of the subject via discourse linking is only possible if the new topic is of a different type, namely an adjunct (6a) but not an argument (6b).

(6) a. *Mingtian*$_t$, [_juede [*t* hui xiayu]]. (_ = John).
Tomorrow, [_ believe [*t* will rain]].
'It is tomorrow that John believes will rain.'
b. * *Bill*$_t$, [_juede [*t* shi laoshi]]. (_ = John).
Bill, [_ believe [*t* is teacher]].
'It is Bill that John believes is the teacher.'

Even more direct parallelism comes from Brazilian Portuguese, a language that has overt movement in wh-questions (like English) but omits arguments in certain contexts (like Chinese).[3]

(7) a. Quando/Como/Onde$_t$ _ beijou *t*?
When/How/Where$_t$ _ kissed$_{2/3P}$ *t*?
'When/How/Where$_t$ did you/they kiss?'
b. * Quem$_t$ _ beijou *t*?
Whom$_t$ _ you/they kiss$_{2/3P}$ *t*?
'Whom did you/they kiss?'

The verbal morphology of Brazilian Portuguese has become too impoverished to support the agreement-based pro-drop option in its European cousin, as can be seen in the inflectional form of *beijou* in (7). Topic drop á la Chinese is the only option, and we see the exact asymmetry between adjunct and argument wh-questions – in child *English*.

Some may object to calling the topic-drop grammar a *parameter* but what's at stake is surely not terminological. The main generalization is that English-learning children spontaneously exercise a grammatical option never attested in their environment but used by speakers thousands of miles away. This option has to be suppressed by language-specific data: namely, the use of expletive subjects such as *There is a car coming* and *It seems that the kids are tired* because they are not thematic,

3. I thank Pablo Faria and Guilherme Garcia for the data reported here.

thereby only serving a formal requirement of the English grammar. For instance, where the grammatical subject *it* must be present for the English expression *It is going to rain*, the position can be empty in Chinese ('_ yao xiayu' or "will rain"; Wang et al., 1992). Expletive subjects needn't be assumed to be a trigger innately associated with the English-type grammar (e.g., Hyams, 1986). All that's required is for children to understand that the expletive subject, which does not receive a thematic role, must be a formal requirement in the grammar; i.e. that the grammatical subject position needs to be overtly filled. Because expletive subject sentences are infrequent, making up about 1% of child-directed input, the rise of the obligatory subject grammar is gradual, according to the variational model. And the topic-drop grammar will be exercised during the process, resulting in null subjects in child English as well as occasionally null objects, when the object happens to be the discourse topic.

Naturally, the same model ought to account for the acquisition of topic- and pro-drop languages (Valian, 1991; Wang et al., 1992; Kim, 2000; Grinstead, 2000) which, in contrast to the considerable delay in English, show very early adult-like command of subject use. I summarize these findings in Section 5.1, where the variational model is extended to L2 acquisition.

3. Rules vs. storage

At the 2005 LSA summer institute, I organized a workshop called "Nuts and Core",[4] borrowing Culicover's (1999) term for linguistic idiosyncracies that cannot all be plausibly attributed to an innate grammatical core (Chomsky, 1981). The questions posed to the participants, all prominent scholars in the generative vs. constructivist debate, were as follows:

(8) a. If the core is dispensed with, how does the learner go from specific constructions to general regularities in syntax (Tomasello, 2003)? What kind of constraints are needed for learning to be efficient and successful?
  b. If the core is to be maintained, how might one construe a principled theory that keeps the core and the nuts separate (Fodor, 2001)? How does the setting of a parameter value tolerate exceptions?

The workshop was lively but ended in a state of impasse. A main point of contention was whether child language is abstract and productive or item-based and lexically conservative. In many ways, the debate resembled the earlier competence-performance dispute: both sides offer useful but only partial explanations of child language. Later in this paper I will turn to my own proposal of how children discover

4. https://linguistlist.org/issues/16/16-2050.html.

productive rules – and potentially a reconciliation of the two approaches – but first, an assessment of the empirical and theoretical claims is in order.

3.1 Assessing usage-based learning

The first point to make is that an early stage of lexically specific language is neither a novel observation nor a feature unique to the usage/construction-based approach (despite being its "central tenet"; Diessel, 2013). Again, let's turn to the study of English past tense. A well-known pattern is the U-shaped developmental curve. Children's verbal inflection is initially conservative: very few regular verbs are consistently marked in past tense, and the irregular verbs, when marked, are marked correctly. This stage is followed by the emergence of overregularization errors, which we can observe in longitudinal records. For instance, Adam's transcripts started at 2;3; all irregular verbs were marked correctly until 2;11, when he produced the utterance "What dat feeled like" (Marcus et al., 1992; Pinker, 1995). Since *feeled* cannot be attributed to the input, the error marks the elevation of "-ed" to the status of a productive suffix. Thus English past tense is a classic case of initial conservatism followed by productive generalization. The phenomenon was central to the past tense debate and especially the dual-route model developed by Clahsen, Marcus, Pinker, and Prince: all avowed nativists.

The second, and more important, point is empirical: the evidence for an initial item-based stage of child language has been overstated. For example, high frequency combinations such as "give me" (sometimes "gimme") have been interpreted as "unanalyzed" collocations and presented as evidence for a lexically specific stage of language development (Lieven et al., 1992; Tomasello, 1992, 2003). These expressions are indeed statistically dominant, but to conclude that they are item based requires more work. At a minimum, one needs to show that their frequencies are conspicuously higher than expected had the verb and the pronoun been combined independently. In this light, consider the transcripts of the Harvard children Adam, Eve, and Sarah (Brown, 1973), datasets that have been in the public domain for decades. Searching for the strings *give me, give him* and *give her* in the three children's production data shows that they appear 95, 15, and 12 times, for a ratio of 7.75:1.23:1. It is thus quite likely that when working with a relatively small child corpus, *give* is only paired with *me* – perhaps the reason behind *give me* as a paradigm case of item-specific learning. But another search shows that the frequencies of *me, him,* and *her* in these children's production data are 2,949, 484, and 375, or 7.86:1.29:1. In other words, the frequency of "give me" actually suggests that the verb and the object combine independently and productively.

Quantitative research in the usage-based learning literature does not seem to have developed, and assessed, a coherently formulated null hypothesis: namely, that child language is productive, or that child language is *not* usage-based, however a usage-based account is formulated (and it is generally vague; see Tomasello, 2000b for suggestions). Consider three key case studies in Tomasello's influential paper *Do young children have adult syntactic competence* (2000a) and other publications that are often cited as evidence for usage-based learning:

(9) a. The Verb Island Hypothesis (Tomasello, 1992). Most of the verbs and predicates in early child language are used with one or very few possible frames.
b. Limited morphological inflection (Pizzuto & Caselli, 1994). Almost half of the verbs in child Italian were used in one person-number agreement form (out of six possibilities), and only 13% of all verbs appeared in four or more forms.
c. Determiner imbalance. Pine and Lieven (1997) find that only 20–40% of the nouns that have been used the determiner *a* or *the* are used with both, despite the general interchangeability of the determiners (e.g., *a/the dog, a/the chair,* etc.).

So far as I can tell, these claims have been presented without evaluating an alternative hypothesis. At a minimum, it would have been worthwhile to subject *adult* language to item-based claims. With respect to determiner use (9c), quantitative analysis of English print materials such as the Brown Corpus (Kučera & Francis, 1967) and child-directed speech reveals comparable, and comparably low, combinatorial diversity as children (Valian et al., 2009), yet adults' grammatical ability is not in question.[5] I now review a statistically rigorous assessment of what constitutes evidence for a productive grammar (Yang 2013a). Usage-based claims can only be established if the alternative, grammar-based, hypothesis can be rejected.

3.2 A statistical benchmark for grammar

To develop a principled quantitative interpretation of language, we must take Zipf's Law (1949) into account. For reasons no one quite understands (Miller, 1957; Chomsky, 1958; Mandelbrot, 1953; see Yang, 2013b for an exposition), word frequency is inversely proportionally related to rank. Specifically, let there be N unique words in a corpus. For the r-th ranked word, its frequency f is C/r where C is some constant. Thus, its probability of use p can be expressed as:

5. Similar observations hold for verb islands (9a) and inflectional morphology (9b); see Kowalski and Yang (2012) and Yang (2016, chapter 2).

(10) $$p = \frac{C/r}{C/1 + C/2 + \ldots + C/N}$$

$$= \frac{1}{rH_N} \text{ where } H_N = \sum_{i=1}^{N} \frac{1}{i} \text{ is the } N\text{th Harmonic number}$$

Particularly useful here is the approximation of the Harmonic number (H_N) as $\ln N$ which considerably simplifies the calculation of the statistical test (and the productivity model presented in Section 4).

The most characteristic feature of Zipf's Law is the long tail: most linguistic units such as words, and by extension, combinations of words, are rarely used even in very large corpora (Jelinek, 1998). This suggests that the sparsity of syntactic combinations in children's early language, or indeed any linguistic sample, is inherent: It does not automatically support lexically specific learning; as I discuss presently, it may even support a rule-based grammar.

The statistical benchmark for grammatical productivity (Yang, 2013a) incorporates Zipf's Law to approximate word probabilities and their combinations. Let's consider its application to the sparsity of determiner-noun combinations in child and adult language. Suppose we have a corpus of N (singular) noun types that appear in S pairs of *a/the*-noun combinations. The expected probability for the r-th ranked noun having been paired with both *a* and *the* is given below:

(11) $$E_r = 1 - (1-p)^s - \sum_{i=1}^{2} \left[(f_i \boldsymbol{p} + 1 - p)^S - (1-p)^S \right] \text{ where } p = \frac{1}{rH_N}$$

Here f_1 and f_2 are the probabilities of the two determiners. The reader is directed to Yang (2013a) for mathematical details but the most important point about the equation in (11) is highlighted in boldface, where we multiply the probability of the noun (p) with those of the determiners (f_1 and f_2). This assumes that their combinations are statistically independent, that is, *not* lexically specific. If a sample of determiner-noun combinations has been generated by an abstract and productive rule, then the average diversity value calculated from (11) should closely match the empirical value, namely the percentage of the N nouns used with both determiners. As shown in Table 1, although the combinatorial diversity is quite low across both child and adult languages, it is statistically indistinguishable, using tests such as the concordance correlation coefficient test (Lin, 1989) from the expected diversity under a rule where the combinations are fully productive.[6]

6. As Table 1 also makes clear, the actual value of diversity does not tell us anything about the underlying productivity of the grammar: unlike the claims in the usage-based literature, a higher diversity value (such as Nina) does not mean a "more" productive rule than a lower diversity value (such as the Brown corpus); see Pine et al. (2013) for a recent example of this

Table 1. Empirical and expected combinatorial diversity in L1 English (adapted from Yang, 2013a)

Subject	Sample size (S)	Types (N)	Empirical	Expected
Naomi (1;1–5;1)	884	349	19.8%	21.8%
Eve (1;6–2;3)	831	283	21.6%	25.4%
Sarah (2;3–5;1)	2453	640	29.2%	28.8%
Adam (2;3–4;10)	3729	780	32.3%	33.7%
Peter (1;4–2;10)	2873	480	40.4%	42.2%
Nina (1;11–3;11)	4542	660	46.7%	45.1%
Brown corpus	20650	4664	25.2%	26.5%

The method developed in (11), which has been independently replicated by other groups (e.g., Silvey & Christodoulopoulos, 2016), has broader applicability, benefiting from the accuracy of Zipf's Law (10), or $p = 1/rH_N$. We can calculate the expected combinatorial diversity based only on the sample size (S) and types (N) appearing in the sample, without even knowing the identities of the words. In recent work (Goldin-Meadow & Yang, 2017), the method has been applied to home signs, the gestural systems created by deaf children with properties akin to grammatical categories, morphology, sentence structures, and semantic relations found in spoken and sign languages (Goldin-Meadow & Mylander, 1998). Quantitative analysis of predicate-argument constructions suggests that, despite the absence of an input model, home signs show the expected degree of combinatorial productivity. By contrast, the test has also been used to provide rigorous supporting evidence that Nim Chimpsky, the chimpanzee raised in an American Sign Language environment, never mastered the productive combination of signs (Terrace et al., 1979; Terrace, 1987): Nim's combinatorial diversity such as "give Nim" and "give me" falls far below the level expected of a productive rule (Yang, 2013a).

I must be clear about what the determiner productivity study does and does not show. It demonstrates that, at least for one aspect of child language, combinatorial productivity is on full display from the earliest testable stage of multi-word combinations. Thus, the usage-based claim for a lexically specific grammar is not supported. Furthermore, it provides a methodological example of how to develop statistically rigorous assessments of language data, a point to which I return in Section 5. But I do not suggest that all aspects of child language are productive from the get-go: see the remarks about English past tense earlier. Even for the determiner system, what we have established is a *formal* aspect of the grammar, that

fallacy. The formal analysis of the variation and how the empirical data are not predicted by usage-based learning models can be found in Yang (2013a) .

(at least) two syntactic categories combine freely and productively. Other properties of the determiner system – for instance, semantic properties such as count and mass, and pragmatic properties such as specificity – may take a good deal of fine tuning: non-target forms such as "a hair", "a blood", "a dirt", etc. are not uncommon in the speech of children whose determiner-noun combinatorial productivity is not in doubt. In fact, the successful acquisition of the formal system of determiner use appears to help establish the count/mass distinction in child language (e.g., Gordon, 1985, 1988). In addition, the English determiner system contains some truly idiosyncratic elements: for example, while (American) English does not use determiners with place names ("*the Chicago", "*the Brooklyn"), there is "the Bronx", which can only be lexically memorized.

The very early syntactic productivity of the determiner system raises important questions: *How* do children learn this particular rule of the English grammar so quickly? Here Zipf's Law poses a significant challenge. The sparsity of language entails that the caretaker's speech can never be fully saturated with linguistic combinations. Recall that the combinatorial diversity in caretaker speech is equally low as in child speech (Valian et al., 2009): that is, only a small fraction of nouns that can combine with both *a* and *the* will do so in the input. So how does the child generalize a property that holds for a small subset of words to all words – as they apparently do in Table 1? To answer this question, we need to confront the problem of learning by generalization: How do children acquire productive rules from lexical examples?

4. Productivity and exceptions

As Sapir remarked "all grammars leak" (1928, p. 38–39): the balancing act between rules and exceptions is one of the oldest problems in linguistics. While linguists can distinguish rules from exceptions by carrying out grammaticality judgments and reaction time experiments (e.g., Clahsen, 1999), children face a more formidable challenge. Since rules and exceptions are defined in opposition of each other, children seem to face a chicken-and-egg problem, and it is one that needs to be resolved in a few short years, without supervision or feedback, all the while under the sparsity of data befitting Zipf's Law.

4.1 Two kinds of exceptions

It is useful to distinguish two kinds of exceptions to rules. One kind can be called *positive exceptions.* The English past tense system is such an example: children receive overt evidence for the exceptions against the rule, by hearing irregular past tense forms that do not take "-ed". This is a familiar problem. A great many approaches, ranging from generative linguistics (e.g., Aronoff, 1976, p. 36) to connectionist modeling (Marchman & Bates, 1994) to hybrid models (Pinker, 1999), share the same underlying intuition: a rule must "earn" its productivity, in the sense that it must somehow overcome the exceptions. It is frequently observed that a productive rule ought to be the one that covers the most diverse range of items: indeed, "statistical predominance" is traditionally the hallmark for linguistic productivity (e.g., Nida, 1949, p. 14). However, (12) provides some illustrative problems from morphology and phonology, which suffice to show that the solution is not so simple:

(12) a. English past tense: A default rule is learned abruptly and results in overregularization, after a protracted stage of rote memorization (Marcus et al., 1992; Yang 2002).
b. English stress: The grammar of English stress (Chomsky & Halle, 1968; Hayes, 1982; Halle & Vergnaud, 1987) is not trochaic with a list of lexical exceptions despite a vast majority of English words bearing stress on the first syllable (Cutler & Carter, 1987; Legate & Yang, 2013).
c. German noun plurals: A suffix ("-s") can be the productive default despite coverage of fewer nouns than any of its four competitors (Clahsen et al., 1992; Wiese, 1996).
d. Russian gaps: Morphological categories need not and sometimes do not have a default, as illustrated by the missing inflections of certain Russian verbs in the 1st person singular non-past (Halle, 1973). Such cases are far from rare (Baerman et al., 2010): the absence of the past tense for *undergo* and past participle for *stride* are the more familiar examples from English speakers (Pullum & Wilson, 1977; Pinker, 1999). For discussion in the context of L1 acquisition, see the curious case of Polish masculine genitives (Dąbrowska, 2001; Yang, 2016).

By contrast, *negative exceptions* in language learning seem more paradoxical. These are the cases where children must learn that a rule does *not* apply across to items that it could have (e.g., "minor rules" of Lakoff, 1970, a long-standing problem). The well-researched English dative constructions illustrate the nature of the problem clearly:

(13) a. John gave the team a prize. John gave a prize to the team.
b. John assigned the students a textbook. John assigned a textbook to the students.
c. *John donated the museum the painting. John donated the painting to the museum.
d. John guaranteed the fans a victory. *John guaranteed a victory to the fans.

The verbs *give* and *promise* can freely alternate between the double object construction and the *to*-dative construction. However, semantically very similar verbs such as *donate* can only appear in the *to*-dative construction, and *guarantee* is exactly the opposite. Because children do not receive negative evidence, how do they learn what not to say in their language? Here a lexically conservative approach cannot work: the productivity of these constructions is evident in child language (Gropen et al., 1989; Conwell & Demuth, 2007) and can also be observed when they are extended to novel verbs with appropriate semantic properties: when the verb *text* appeared, its double object form was instantly available as in *I texted them the score.*

Problems such as the acquisition of the dative constructions once dominated the learnability research in the generative tradition (Baker, 1979; Berwick, 1985; Fodor & Crain, 1987; Pinker, 1989). In recent years, they have become a major focus of usage-based theories under the tenet of entrenchment (Tomasello, 2003; Bybee, 2006; Ambridge et al., 2008) or preemption (Stefanowitsch, 2008; Boyd & Goldberg, 2011): both are a form of indirect negative evidence which takes the absence of evidence as evidence of absence (Pinker, 1989). According to a recent formulation (Ibbotson & Tomasello, 2016), "if children hear quite often *She donated some books to the library,* then this usage preempts the temptation to say *She donated the library some books.*" But use of indirect negative evidence is problematic (Pinker, 1989): the absence of evidence is *not* evidence of absence. As I have argued elsewhere (Yang 2015b, 2017) using realistic child-directed data, it is generally impossible to distinguish ungrammatical sentences, which would never appear in the input, from grammatical ones which just happen not to be sampled – because the space of linguistic combinations is enormous, and their statistical distribution is inherently sparse. More empirically, such approaches fail to account for some of the most robustly attested errors in child language such as "I said her no" (Bowerman, 1982; Bowerman & Croft, 2008). The communication verb *say,* of course, is always used in the *to*-dative construction, and is among the most frequently used verbs in English – yet this (deeply) entrenched form fails to preempt the double object construction.

4.2 The Tolerance Principle

The Tolerance Principle and its corollary the Sufficiency Principle (Yang, 2016) provide a unified solution for the problem of rules and exceptions. I will not review the empirical motivation for their development but will simply state:

(14) a. *Tolerance Principle*
Suppose a rule R is applicable to N items in a learner's vocabulary, of which e items do not follow R and are thus exceptions. The necessary and sufficient condition for the productivity of R is:

$$e \leq \theta_N \text{ where } \theta_N := \frac{N}{\ln N}$$

b. *Sufficiency Principle*
Suppose a rule R is applicable to N items in a learner's vocabulary, of which M follow R and no information is available about the remaining $(N - M)$ items. The necessary and sufficient condition for the productivity of R is:

$$(N - M) \leq \theta_N \text{ where } \theta_N := \frac{N}{\ln N}$$

The unifying theme of the two principles lies in the quantity of positive evidence – $(N - \theta_N)$ in both cases – that is necessary to support the productivity of a rule. To understand the intuition behind the rationale, consider an analogous case in a non-linguistic domain. Suppose you have encountered 10 new species off a remote island, of which 8 share a certain property (e.g., phosphorescence). Even though you may not have any information about the other two, or maybe even if the other two are known *not* to have the property, it seems reasonable to form a generalization about the entire class and extend it to the 11th species. By contrast, if the property only holds for 2 of the 10 examples, it seems wise not to rush to any general conclusion. The Tolerance Principle provides a precise weight of evidence, in the form of $\theta_N = N/\ln N$, that warrants productive generalizations.

The mechanism of learning under the Tolerance Principle is schematically illustrated in Figure 1.

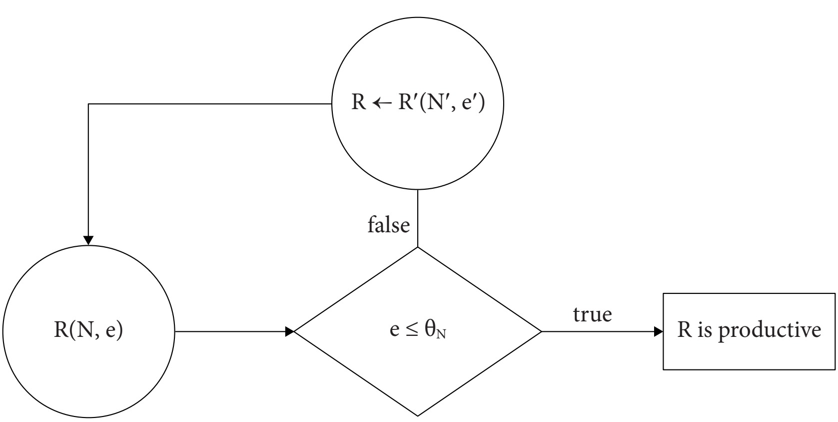

Figure 1. Tolerance Principle guides the search for productive rules in language learning

Language learning is a search for productive generalizations that is best characterized as abduc-tive learning (Chomsky, 1968, p. 80). Children construct a rule *R* from the input data guided by linguistic and cognitive constraints and evaluate its productivity according to the associated numerical values (*N* and *e*). The rule is deemed productive if the positive evidence is sufficiently high; otherwise learners formulate a revised rule (R') to obtain a new set of values (N' and e') and the Tolerance/Sufficiency Principle is applied recursively. Thus, the quantitative accumulation of exceptions can lead to the qualitative change in the productivity of rules. An illustrative case study can be found in the analysis of English stress rules referred to in (12). When the child reaches a modest vocabulary (i.e., *N* is relatively large), the rule that stresses the first syllable fails to meet the productivity threshold despite covering the majority of words (Cutler & Carter, 1987). The child subsequently divides the vocabulary into nouns and verbs, and productive rules within each subcategory can be established within (Legate & Yang, 2013; Yang, 2016); I will return to this theme in the context of multilingual acquisition in the final remarks of this paper. If no plausible rule can be found that meets the criterion for productivity, children will lexically memorize each instance that follows *R* and no productive generalization will be established.

Table 2 provides some sample values of *N* and the associate threshold values θ_N, the maximum number of exceptions that a productive rule can tolerate:

Table 2. The maximum number of exceptions for a productive rule over *N* items

N	θ_n	%
10	4	40.0
20	7	35.0
50	13	26.0
100	22	22.0
200	38	19.0
500	80	16.0
1,000	145	14.5
5,000	587	11.7

These thresholds for productivity under the Tolerance Principle are significantly lower than a naïve "majority rule", which has many interesting implications for language acquisition. In particular, the Tolerance Principle asserts that a smaller vocabulary (i.e., smaller values of *N*) can tolerate a higher percentage of exceptions: all else being equal, productive rules are *easier* to detect for learners who have access to *less* input data. I return to this important theme in the conclusion.

Here I briefly summarize the application of the Tolerance and Sufficiency Principle to two of the most intensively studied problems in language acquisition: the past tense and the dative constructions in English.

In the case of English past tense, a distributional learning of induction must be used to identify the rules of verbal infection. Many proposals of inductive learning, from many diverse fields (e.g., Chomsky, 1955; Osherson & Smith, 1981; Mitchell, 1982; Cohen, 1995; Yip & Sussman, 1997), are applicable. These models typically operate by forming generalizations over exemplars in some suitable representation. For example, suppose two good baseball hitters can be described with feature bundles [+red cap, +black shirt, +long socks] and [+red cap, +black shirt, +short socks]. The rule "[+red cap, +black shirt] → good hitter" will follow, as the shared features (cap, shirt) are retained and the conflicting feature (sock) is neutralized. This method is thus very capable of identifying the "-ed" suffix as one applicable without all phonological restrictions on the stem: the verbs that take "-ed" are phonologically very diverse, and no restrictions will be identified (Yip & Sussman, 1997). Thus, the productivity of "-ed" will be determined by the total number of verbs (N) and the number of irregular exceptions (e) in the learner's vocabulary. The same consideration must also apply to the patterns that govern irregular verbs. For instance, the irregular verbs *bring, buy, catch, fight, seek, teach,* and *think* all undergo a stem change replacing the rime with [ɔt]. This rule also has no restriction on the stem, because the participating verbs are phonologically very diverse. But it is easy to see that the rule "rime → ɔt" will fare terribly: the seven positive members are easily swamped by hundreds of negative examples that do not change the rime to [ɔt], far exceeding the tolerance threshold. As a result, the rule is not productive and will be lexicalized – to these seven verbs. Other irregular patterns can be analyzed similarly: as shown elsewhere (Yang, 2016, chapter 4), all rules except the regular "add *-d*" will be assessed as unproductive, accounting for the near-total absence of over-*ir*regularization errors in child English (Xu & Pinker, 1995; see Lignos & Yang, 2016 for a cross-linguistic review of similar findings.)

Following the same logic, we can see that the emergence of the "add *-d*" rule will require a long gestation period. Although children can quickly induce its structural description – perhaps using no more than a few dozen verbs (e.g. Yip & Sussman, 1997) – irregulars are likely overrep resented in children's early vocabulary. For instance, in a 5-million-word corpus of child-directed English extracted from the CHILDES database (MacWhinney, 2000), 76 of the 200 most frequent verbs in past tense are irregular. As θ_{200} is only 37, children with a small vocabulary are unlikely to establish the productivity of "-ed", despite the fact that it may be by far the statistically dominant rule. For very young children, then, verbs marked with "-ed" are in effect on par with irregulars: they are item based and lexically memorized.

Telltale evidence for productivity comes from the first attested overregularization errors. When longitudinal records are available, the Tolerance Principle can account for the particular juncture at which productivity emerges. As noted earlier, Adam produced his first recorded overregularization error at 2;11 ("What dat feeled like?"). By then, Adam must have acquired a sufficiently large number of regular verbs to overwhelm the irregulars. In Adam's transcripts leading up to 2;11, he used $N = 300$ unique verb stems in all, of which $e = 57$ are irregular. This is quite close to the predicted $\theta_{300} = 53$, and the discrepancy may be due to the underrepresentation of his regular verbs, which will be less frequent and thus more likely to be left out of a sample. Thus, Adam acquired a productive "-ed" only after he acquired a "super" majority of regular verbs, consistent with the Tolerance Principle.

4.3 Dative generalizations and retreats

The acquisition of the dative constructions is more complex as it involves children overgeneralizing before retreating: young children say "I said her no" but older children and adults do not. But the same procedure of constructing and evaluating generalizations applies here as well. Here I will briefly review the acquisition of the double object frame (V NP NP) as the *to*-dative construction can be handled in a similar fashion (Yang, 2016, chapter 6). In a five-million-word corpus of child-directed English data, roughly corresponding to one year worth of input, we can find a total of 42 verbs used in the double object construction. Of these, 38 have a very clear semantics of "caused possession", which we assume is identifiable if the learner is equipped with a suitable set of conceptual and semantic primitives (e.g., Pinker, 1989; Grimshaw, 1990; Jackendoff, 1990). The four exceptions, well below the threshold $\theta_{42} = 11$, do not convey caused possession: all are performative verbs (call, *consider, name,* and *pronounce,* e.g., *I called him a liar).* Thus, the semantic condition necessary for double object construction (Gropen et al., 1989; Pinker, 1989; Levin, 1993; Goldberg, 1995; Pesetsky, 1995; Krifka, 1999) need not be stated as a UG primitive but can acquired from the language-specific data. The following hypothesis, then, can be formulated under the guidance of the Tolerance Principle:

(15) If a verb appears in the double object construction, then it will have the semantics of caused possession.

At this point, the child may consider the converse of (15), in trying to establish the validity of caused possession as a *sufficient* condition for the double object condition. This amounts to testing if the entire set of caused-possession verbs (N) in the child-directed corpus can be used in the double object construction. According to the Sufficiency Principle, this is warranted only if the $M = 38$ items, which are actually attested in the construction, constitute a sufficiently large subset of N.

In the present case, the input corpus contains an additional 11 verbs that belong to the semantic class in (15). But these did not appear in the double object construction:

(16) address, deliver, describe, explain, introduce, return, transport, ship, mention, report, say

(16) is an interesting list. For some of the items (e.g., *deliver* and *say*), the double object construction is ungrammatical: **John delivered the kids a pizza, *John said Bill something mean.* But the verb *ship* does allow the double object construction – *John shipped Bill his purchase* – it just was not in the corpus because the caretakers opted for the to-dative form instead. Of course, the child does not know *why* the 11 verbs in (16) fail to show. The statistical distribution of language makes it difficult, if not impossible, to distinguish impossible forms from possible but unattested forms, which is the Achilles heel of indirect negative evidence and its entrechment or preemption variant (Yang, 2015b, 2017). Nevertheless, $M = 38$ constitutes a sufficiently large subset of $N = 49$ since $49/\ln 49 = 12$, and thus the following generalization ensues:

(17) If a verb has the semantics of caused possession, then it can appear in the double object construction.

This immediately accounts for the overgeneralization errors such as *I said her no.* A search of child speech data in CHILDES also yields errors such as *I delivered you a lot of pizzas* (3;8), a verb not permissible in the construction in the adult language, thereby supporting the productivity of (17). This also accounts for the experimental evidence that children as young as 3;0 have productive usage of the dative constructions upon learning a novel verb with the appropriate semantic properties (e.g., Gropen et al., 1989; Conwell & Demuth, 2007).

The retreat from overgeneralization straightforwardly follows the Tolerance/Sufficiency Principle but to do so requires the child to expand their vocabulary. Highly frequent verbs, which populate our child-directed corpus and are among those learned earlier by children, heavily favor the productive use of the rule in (17). I combined several corpora to approximate the vocabulary of ditransitive verbs likely known to most English speakers (Yang, 2016, p. 208), and the results are given in Table 3.

Table 3 shows that a child with a very limited vocabulary is bound to conjecture (17) as a productive rule, namely, *all* caused-possession verbs may appear in the double object construction. But as they learn more verbs, the proportion of those used in the construction will continue to drop, eventually below the sufficiency threshold. The child can thus successfully retreat, without the problematic use of indirect negative evidence.

Table 3. Caused-possession verbs and their availability in the double object construction (adapted from Yang, 2016)

top	yes	no	θ_N	productive?
10	9	1	4	Yes
20	17	3	7	Yes
30	26	4	9	Yes
40	30	10	11	Yes
50	34	16	13	No
60	39	21	15	No
70	43	27	16	No
80	46	24	18	No
92	50	42	20	No

It is interesting to probe the properties of the verbs in Table 3 further, which provides a learning-theoretic account for many regularities in the double object construction and its acquisition (Mazurkewich & White, 1984; Gropen et al., 1989; Bley-Vroman & Yoshinaga, 1992; Inagaki, 1997); see Yang and Montrul (2017) for a review. For example, 50 of the 92 verbs in Table 3 are monosyllabic, of which 42 allow double objects. By comparison, only 10 of the 42 polysyllabic verbs can participate in the construction. Thus, when given a novel verb that describes, say, the movement from an object initiated by one individual to another, both children and adults are more inclined to accept a short novel verb (e.g., *pell*) in the double object construction than a long one (e.g., *orgulate*). These tendencies are simply the consequences of distributional learning, rather than structural constraints as proposed in some theoretical literature (e.g., Harley & Miyagawa, 2016). Similarly, while the entire class of caused-possession verbs cannot categorically participate in the double object construction, productivity can be found in semantic subclasses assuming that such classes can be constructed by language learners (Pinker, 1989; Yang, 2016): recall that productivity is more likely with smaller values of *N*.

In sum, the Tolerance Principle appears to embody what some usage-based researchers envision as the key solution to the problem of language learning: "a single mechanism responsible both for generalization, and for restricting these generalizations to items with particular semantic, pragmatic, phonological (and no doubt other) properties (Ambridge & Lieven, 2011, p. 267)". Its simplicity yields sharp behavioral predictions, with interesting implications for the apparent differences between the outcome of L1 and L2 acquisition.

5. Formal applications to L2 acquisition

The mathematical models reviewed here address the type of questions that arise for language acquisition quite generally – by children and adults alike. Here I offer some comments and speculations on how these methods may apply to adult language acquisition.

5.1 UG access and input frequency

Any learning model must consist of precise statements about the hypothesis space that the learner entertains – the initial state – and the mechanisms of learning from data (Chomsky, 1965; Yang, 2002). A successful model is the combination of the initial stage and the learning mechanisms that explains the specific patterns of language development. These questions are often debated in L1 acquisition but they arise for L2 as well, often in the form of the role concerning UG (e.g., Clahsen & Muysken, 1986; Cook & Newson, 2014; Epstein et al., 1996; Schwartz & Sprouse, 1996; White, 2003; Rothman & Slabakova, 2017). Adult language learners also have an initial state, which may be partly dependent on their first language, and they also need to integrate data and experience into their linguistic knowledge. I would like to suggest that the variational learning model provides a useful perspective on these theoretical issues.

It is now evident that language, and language learning, obey certain structural constraints: not all formal language-like systems are natural or naturally learnable for by otherwise capable children and adults alike (Crain & Nakayama, 1987; Smith & Tsimpli, 1995; Newport & Aslin, 2004; Tettamanti et al., 2004; Friederici, 2017). What remains controversial is the nature of such constraints: what they are and whether they reflect domain specific restrictions on language or follow from more general cognitive principles. If it can be established that the variational model is employed by child and adult language learners alike, essentially by holding the learning mechanism component of the model constant, then the nature of the initial state may be fruitfully investigated.

The variational learning model is very likely implicated in language acquisition throughout the lifespan. In fact, it would be highly surprising if it were not, given its ubiquity in probabilistic learning and decision making across domains and species (Estes, 1950; Bush & Mosteller, 1951; Herrnstein & Loveland, 1975), including language (Labov, 1995; Roberts, 1997; Hudson Kam & Newport, 2005; Smith et al., 2009; Miller & Schmitt, 2012). And there appear to be strong continuities between children and adults in at least certain domains of language. For word learning, the parallels between children and adults are very strong (Markson & Bloom, 1997; Bloom, 2000), and models of lexical acquisition are frequently tested on adult participants (e.g., Yu & Smith, 2007; Stevens et al., 2016). Furthermore, the model converges to a statistical combination of multiple hypotheses in linguistically heterogeneous environments as in the case of bilingualism and language change (Yang, 2000), which further supports its broad applicability (see, e.g., Slabakova, 2008). Although there is no denying that adult, non-native grammatical acquisition is different from that of children in path and ultimate attainment, the differences do not have to relate to differences in underlying (cognitive/linguistic) mechanisms available to each. Let us assume, then, that variational learning is indeed used by language learners at all stages of development and explore how it helps determine the role of UG in adult language acquisition.

Consider, again, the obligatory use of grammatical subjects in languages such as English. As reviewed earlier, L1 acquisition can be modeled as a competition among the options delimited by UG: the telltale evidence is frequent occurrence of null subjects in adjunct Wh-questions (3) and near absence in argument Wh-questions (4), both characteristic of a topic-drop grammar. How would an adult learner acquire a second language which uses the grammatical subject in a different way from their their first language?

In (18), I summarize the distributional evidence, extensively discussed elsewhere (Yang, 2002), that can disambiguate the three broad classes of grammars from the perspective of the language learner. Their attested frequencies, as a percentage of the utterances in child-directed Chinese, Italian, and English are also reported.

(18) a. Chinese: Null objects (11.6%; Wang et al., 1992)
b. Italian: Null subjects in object wh-questions (10%; Yang 2002)
c. English: Non-referential expletive subjects (1.2%; Yang 2002)

The amount of unambiguous evidence for the Chinese and Italian-type grammars is quite high. It is in fact, higher than the unambiguous evidence – about 7% (Yang, 2002) – for the correct placement of the finite verb, which children acquire very early on and are essentially error-free (Pierce, 1992). Thus, we predict very early acquisition of topic drop and pro drop. Indeed, studies of the acquisition of Italian (Valian, 1991), Catalan/Spanish (Grinstead, 2000), Chinese (Wang et al., 1992), Korean (Kim, 2000), etc. have consistently found that children reach adult-level use of subjects around age 2, considerably earlier than the consistent use of grammatical subjects by English-learning children which typically takes place by age 3 or later.

Note that these findings about child language only follow if the initial stage consists of hypotheses, and their associated properties, laid out in (18). Because virtually all grammatical subjects are also thematic, the protracted null subject stage in child English must be linked to the low frequency of expletive subjects, which are the only type of input that distinguishes the competing options in UG. Suppose the initial state of the grammar is a probabilistic form of phrase structure rules such as $S \xrightarrow{p} NP\ VP$, where p would be nearly 1.0 for English because the subject is almost always present and would be somewhere around 0.5 for Chinese (Wang et al., 1992), which allows topic drop. It is easy to see that if such rules were used, an English-learning children should quickly converge to the target grammar, contrary to the protracted subject drop stage in actual language acquisition.

With this background on L1 acquisition, we can turn to the question of adult second language acquisition. In fact, the published literature already points to strong parallels between children and adults: the acquisition of pro drop and topic drop appears "easier" than the acquisition of the obligatory subject use. For instance, Phinney's (1987) classic study finds that while advanced L1 Spanish learners of English do not consistently use expletive subjects, even beginner L1 English learners of Spanish show excellent command of pro drop; see also the later work of Pérez-Leroux and Glass (1999). The topic-drop option is likewise easily acquired. For example, Kanno (1997) finds that L1 English learners of Japanese have close-to-native command of null subjects across a number of syntactic and discourse contexts. By contrast, even near-native L2 learners of English fail to consistently use the expletive subject (Judy, 2011), the true hallmark of the obligatory subject grammar. The variational learning model provides a straightforward account for these cross-linguistic findings. As shown in (18), the advantage of the topic/pro-drop grammars over the obligatory subject grammar is afforded by the more abundant disambiguating evidence in the input language, because variational learning is gradual, probabilistic, and quantity sensitive.

The variational approach to L2 acquisition can help make an even stronger claim about the role of UG. If adult language acquisition mirrors child language acquisition, then one must conclude that the parametric options of UG are available to children and adults alike, which amounts to a very strong form of UG access (White, 1989; Schwartz & Sprouse, 1994, 1996; Epstein et al., 1996; White, 2003). To establish this claim requires the same kind of evidence from child language acquisition under the variational model: we need to find the footprints of non-target yet UG-consistent grammatical hypotheses. We thus turn again to the argument vs. adjunct asymmetry: null subjects are possible in adjunct wh-questions but not possible in argument wh-questions.

Such an asymmetry, if confirmed, would not be very surprising for an L2 learner whose L1 is a topic-drop grammar and transfers into L2, but it would provide strong evidence for the full accessibility of UG for learners whose L1 is a pro-drop language, which licenses null subjects by agreement and does not show this restriction. I am not aware of any study of adult acquisition of the English grammatical subject that specifically targets these distributional properties. In what follows, I can only offer a preliminary analysis based on a corpus provided by Klein and Perdue (Perdue, 1993, available at talkbank.org).

There are four adult English learners whose L1 is Italian, which provide suitable opportunity to examine the distribution of null subjects in their wh-questions. I extracted all of their wh-fronted questions. There are 35 (object) argument questions, only one missing the subject ("what doing in here"). There are 72 adjunct questions, with 16 missing the subject; some examples are given below:

(19) where is?
why shouldn't?
how much cost?
why no have appuntamento (appointment) in the evening?
when come in the school.

The sample size is small but there is a statistically significant difference ($p = 0.01$) between the null subject rates for argument and adjunct wh-questions. The predictions here are straightforward and can be easily verified in future research with larger datasets. The findings are suggestive: adult learners have access to a UG option – topic drop á la Chinese – that is neither in their first (Italian) or second (English) language.

The continuity between child and adult language, if true, points to potential intervention strategies. Although second language learners of English are frequently reminded of the fact that English requires the subject, only expletive subjects truly serve the purpose of driving the learner toward the target form, much like the acquisition of English by young children. Because expletive subjects are relatively infrequent in language use, amplifying the amount of such input may result in accelerated acquisition of the subject, similar to the improvement in L2 speech perception and production under targeted input (e.g., Bradlow et al., 1999). In addition, different intervention strategies may be designed to probe the nature of adult learners' initial state. For example, if adult learners more rapidly acquire the obligatory use of English subjects on the basis of expletive subjects rather than merely lexical or pronominal subjects as in the rule $S \rightarrow$ NP VP, it would provide further evidence for the continuity between child and adult language acquisition and amplify the role of UG.

5.2 Rule, productivity, and vocabulary

The statistical test for assessing grammatical productivity can be ported straightforwardly to the study of second language: Do L2 speakers go through a stage where the grammar is lexically specific and lacks abstract generalization? Claims from the L1 study of usage/item-based learning Tomasello, 2000a, 2003) appear to have been extended into second language research (Ellis & Larsen-Freeman, 2009): low frequency and diversity of syntactic combinations are taken as evidence for lexically based prototypes and exemplars and the absence of a fully abstract system. As discussed in Section 3, these claims cannot be taken at face value unless they are evaluated against rigorously formulated statistical hypotheses, including the null hypothesis that language – of children and adults – is in fact fully productive.

In this section, I apply the methods developed by assessing child language to adult language. The inferential problem is the same: given a linguistic corpus, what is the nature of the underlying mechanism that generates the production? Is it a fully productive system, or is it lexically specific? The case study focuses on seven adult learners of English studied by Klein and Perdue (Perdue, 1993, available at talkbank.org). The methods are very simple and can be automated with simple natural language processing tools (see Yang, 2013a and Silvey & Christodoulopoulos, 2016 for details).

Table 4. Empirical and expected combinatorial diversity in L2 English (data from talkbank)

Subject	L1	Sample size (S)	Types(N)	Empirical	Expected
Andrea	Italian	355	171	4.1%	9.8%
Lavinia	Italian	822	295	19.3%	22.6%
Santo	Italian	398	193	3.1%	9.6%
Vito	Italian	314	139	6.5%	11.0%
Ravinder	Punjabi	121	75	5.3%	8.6%
Jainail	Punjabi	283	148	8.1%	9.3%
Madan	Punjabi	237	102	3.9%	11.7%

The L2 learners' use of determiner-noun combinations is significantly below the diversity level expected under a fully productive grammatical rule ($p < 0.001$; paired one-tailed Mann-Whitney test). Note however the samples in Table 4 are considerably smaller than the datasets from L1 acquisition and the results must be taken with a grain of salt.[7] But Table 4 does appear to reveal a usage-based stage of L2 acquisition in which learners have not mastered a simple grammatical rule, which L1 learners command with ease at a very early age (Table 1). This conclusion is consistent with reports of protracted development of the determiner system in L2 acquisition (Liu & Gleason, 2002; Ionin et al., 2008; Snape, 2008).

To understand the discrepancies between L1 and L2 acquisition, we must first address how young children acquire the determiner-noun rule in the first place.

7. This is also a plea for wider dissemination of L2 data for public use.

Because the rule is language specific, it must be learned distributionally from the input data. An examination of the child-directed input data turns up an interesting puzzle – one which may shed light on why adult language learners struggle with grammatical rules.

Consider Adam, for the last time. He produced 3,729 determiner-noun combinations in his speech with 780 distinct nouns. Of these, only 32.2% appeared with both determiners, which is similar to the expected value of 33.7%; see Table 1. Adam's mother, whose speech was also transcribed in the same corpus, produced a diversity measure of 30.3% out of 914 nouns.[8] Even among the 469 nouns used at least twice, which provided opportunities to be used with both determiners, only slightly over half of them (260) did so. To appreciate the logic of learning, consider a baseball analogy: the interchangeablity of *a* and *the* for a noun can be viewed a batter's ability to switch hit (i.e., batting both left- and right-handed). Suppose a scout has been sent to evaluate a team of players, about a third of whom switch-hitted in the batting practice. It would seem crazy to conclude *all* players in the squad are switch-hitters, but that is apparently what Adam did: he generalized the interchangeability of *a* and *the* from a third of nouns to all nouns. Such an inductive leap seems absurd. It is certainly not sanctioned by the Tolerance/Sufficiency Principle, which asserts that a rule – the interchangeability of *a* and *the* – can be extended to a class of words only if the rule holds for an overwhelming majority of the words.

A promising, and perhaps the only, way out of this dilemma is to make use of a key property of the Tolerance Principle: rule learning is easier, and more tolerant of exceptions, when the learner has a smaller set of items in their vocabulary (Table 2). The developmental literature offers the idea of "less is more" (Newport, 1990; Elman, 1993; Kareev, 1995; Cochran et al., 1999): the maturational constraints place a limit on the processing capacity of young children, which may turn out to be beneficial for language acquisition. If children's vocabulary is smaller, the odds of acquiring productive rules improve considerably.

Consider again the determiner-noun combinations produced by Adam's mother. Only 277 out of the 914 nouns are used with both *a* and *the*, which is nowhere near the requisite threshold for generalization ($\theta_{914} = 134$). But if Adam were only to learn from the 50 most frequent nouns, he would notice that almost all of them – 43 to be precise – are paired with both determiners. On this much smaller subset of data where $N = 50$, there is sufficient evidence for generalization: the 7 nouns that appear exclusively with only one determiner are below the tolerance threshold $\theta_{50} = 12$. For the top $N = 100$ nouns, 83 are paired with both determiners: the 17 loners are again below the tolerance threshold $\theta_{100} = 23$. At the

8. Once again, this low diversity does *not* imply that Adam's mother does not use the determiner rule productively.

time when children acquire the productive rule for determiners, their vocabulary size does not exceed a few hundred (Fenson et al., 1994; Hart & Risley, 1995). It is thus highly likely that they have acquired the rule on a very small set of high frequency nouns, almost all of which will show interchangeability with both determiners; the rest is just noise.

This line of thinking naturally leads us to speculate why adults tend to be worse at language learning than children, when they are better at pretty much everything else. There are of course many differences between children and adults but I would put forward a simple but bold possibility. By the virtue of having greater cognitive capacities, or perhaps due to the influence of their L1 lexicon in developing an L2 lexicon (Jiang, 2000; Dijkstra, 2005; Van Assche et al., 2012; Tokowicz, 2014), adult learners may have know too many words for their own good.

Table 5 gives the token/type ratio for the words used by the L1 and L2 learners, whose productivity measures have been given in Tables 1 and 4. The ratio provides a rough measure of the language user's vocabulary. A ratio of X means that, on average, the speaker produces a new word type every X words; thus, a small token/type ratio is an indication of a larger vocabulary, a long-standing practice in language research (Miller, 1981; Huttenlocher et al., 1991). It is evident that the adults have considerably larger vocabularies than the children.[9]

Table 5. Vocabulary size estimates and productivity in L1 and L2

Subject	Token	Type	Token/type
Adam	140,793	3,811	36.94
Eve	27,147	1,582	17.16
Naomi	35,459	2,274	15.59
Nina	87,933	2,553	34.44
Peter	62,006	1,936	32.03
Sarah	73,951	3,658	20.22
Andrea	8,097	1,013	7.99
Lavinia	16,941	1,673	10.13
Santo	10,705	1,284	8.34
Vito	9,344	1,296	7.21
Ravinder	9,980	954	10.46
Jainail	10,017	913	10.97
Madan	8,672	909	9.54

9. The data is scanty but even at the early testing sessions shortly after their arrival (Perdue, 1993), the adult learners' token/type ratios are still considerably lower than young children's.

I suggest that the L2 learners' apparent inability to use a simple rule fully productively is because they know *too many* words. Young children have no choice but to learn from a small set of high frequency words for which the evidence for productive rules is sufficiently strong. Again, toddlers' vocabulary size has been estimated to be no more than just over a thousand (Huttenlocher et al., 1991; Hart & Risley, 1995), yet children have perfect command of a productive system including the determiner rule reviewed earlier and many aspects of morphology and syntax (see Guasti, 2004 for a review). Thus, a small vocabulary must be sufficient and, if I am correct, necessary, for the acquisition of the essential components of language. Adults, by contrast, are in fact handicapped by their considerably larger vocabulary size due to their mature cognitive capacities. A larger value of N has the inadvertent consequence of raising the threshold for productivity, thereby making rule learning much more difficult.

Finally, a word about the application of the Tolerance Principle in adult language acquisition. The Principle is a method by which the learner evaluates potentially productive hypotheses about language. That these hypotheses are abstractions from input data and are subsequently evaluated against (further) input data entails that a quantitative measure of the input data is absolutely crucial. At this point, it is worth emphasizing that the values of N and e pertain to the vocabulary composition of specific learners, which necessarily vary on an individual basis. Thus, some learners may discover productive rules before others, and there is also the possibility that the terminal state of individuals' grammars varies with respect to the productivity of certain rules; see Yang (2016, chapter 4) for case studies. Obtaining precise quantitative measures of the input data is obviously much harder for adult language learners. But it may be easier to obtain more accurate offline vocabulary measurements for adults, who are likely to be more cooperative than toddlers. Ultimately it is the individual's *internalized* vocabulary that determiners the productivity of rules. Furthermore, it is possible to devise experiments where one can have precise control over the individual's vocabulary with the use of artificial language (Schuler et al., 2016; Schuler, 2017).

5.3 Final remarks

The mathematical models reviewed here, especially the variational learning model and the Tolerance Principle, are meant to complement each other. In this concluding section, I will first describe how these learning mechanisms interact in both L1 and L2 acquisition before making some general methodological remarks.

While making use of domain-general learning mechanisms, the variational model was developed in the "orthodox" framework of parameter setting, where the grammatical options are made available by an innate UG and subject to competition. Later developments in linguistic theories (Chomsky, 2001, 2005; Berwick & Chomsky, 2016) have aimed to reduce the principles specific to language, and the Tolerance Principle can be viewed as a move into that direction: what can be learned abductively from the linguistic data needn't be built into UG.

As discussed throughout the paper, I believe that at the present stage of understanding, certain grammatical options in language concerning the subject still appear unlearned. But as illustrated in the acquisition of the dative constructions (Section 4.3) and the determiner-noun rule in child English (Section 5.2), the Tolerance Principle offers an alternative account of word order phenomena that traditionally fall under the purvey of syntactic parameters such as head directionality. For example, that English prepositional phrases are head initial can be learned if the child keeps track of the positional relation between prepositions – a finite number of items – and their complements and forms a categorical generalization. Even a (suitably) low degree of exceptions can be tolerated. For instance, the vast majority of the verbs in English do not raise in question formation; those that do, namely a small number of auxiliary verbs, can easily be learned as a set distinct from the main verbs.

In some cases, the Tolerance Principle may not be able to identify a single productive rule for the totality of the input data and recursive applications will be required. The most radical case will be bilingual acquisition or transitional stages of language change that has been characterized as "grammar competition" (Kroch, 1989). As suggested elsewhere (Yang, 2016, p. 137), the problem of distinguishing that there are multiple languages in the input, a very first step in multilingual acquisition, is formally equivalent to the problem of distinguishing sub-regularities in a single language. A handful of strange words or visiting relatives with a silly accent will not disrupt the acquisition of a single linguistic system, just as a few irregular verbs do not undermine the regular "add -d" rule. But if the learner's environment consists of significant quantities of multilingual data, then no single language is likely to tolerate the others as exceptions. The learner will be compelled to partition the input into distinct subsystems and develop independent grammars for each, much like the acquisition of the English stress system, where the failure to establish a single productive rule for the entire vocabulary compels children to subdivide words into nouns and verbs and identify distinct stress rules within each class (Legate & Yang, 2013; Yang, 2016). Once multiple grammars are inductively established for a (sufficiently) heterogeneous body of language, the variational model can then apply to determine the course of their competition (e.g., Yang, 2000). This new way of probing the limit of experience and principles not specific to language – the second and third factor in the sense of (Chomsky, 2005) – is a long term project and can be only be successful when we successfully account for the full range of linguistic facts, which traditionally fell under the first factor, Universal Grammar.

To conclude, it is useful to recall that the rise of modern linguistics and cognitive science was enabled by the formal methods introduced by generative grammar: abstraction and idealization over complex phenomena, followed by deductive analysis of nontrivial depth. Sixty years later, the worst one could wish for are proposals where all conceivable factors are thrown into a stew just so no correlation could ever be missed: that would be to return to the Dark Ages of standard social and behavioral sciences. These approaches do not enlighten but only obfuscate (Yang, 2015a).

Much more research, and especially more data, will be needed to further test these mathematical models for L1 acquisition and to verify their applicability to L2 acquisition. All the same, I hope to have conveyed the importance of formal methods to the study of language acquisition. The equations may turn out to be wrong. But one of the most appealing aspects about language is that it is tractable, and even mechanical. We can make progress even by making mistakes so long as the mistakes are precisely formulated.

Acknowledgements

I would like to thank Holger Hopp, Tanja Kupisch, Jason Rothman, Roumyana Slabakova, Neal Snape, and two anonymous reviewers for helpful comments on an earlier draft.

References

Ambridge, B. & Lieven, E. V. (2011). *Child language acquisition: Contrasting theoretical approaches*. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511975073

Ambridge, B., Pine, J. M., Rowland, C. F., & Young, C. R. (2008). The effect of verb semantic class and verb frequency (entrenchment) on children's and adults' graded judgements of argument-structure overgeneralization errors. *Cognition*, 106(1), 87–129. https://doi.org/10.1016/j.cognition.2006.12.015

Angluin, D. (1980). Inductive inference of formal languages from positive data. *Information and Control*, 45(2):117–135. https://doi.org/10.1016/S0019-9958(80)90285-5

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

Baerman, M., Corbett, G. G., & Brown, D. (Eds.) (2010). *Defective paradigms: Missing forms and what they tell us*. Oxford: Oxford University Press. https://doi.org/10.5871/bacad/9780197264607.001.0001

Baker, C. L. (1979). Syntactic theory and the projection problem. *Linguistic Inquiry*, 10(4), 533–581.

Bates, E. & Elman, J. (1996). Learning rediscovered. *Science*, 274(5294), 1849. https://doi.org/10.1126/science.274.5294.1849

Berwick, R. (1985). *The acquisition of syntactic knowledge*. Cambridge, MA: The MIT Press.

Berwick, R. C. & Chomsky, N. (2016). *Why only us: Language and evolution*. Cambridge, MA: The MIT Press.

Bley-Vroman, R. (1989). What is the logical problem of foreign language learning? *Linguistic Perspectives on Second Language Acquisition*, 4, 1–68.

Bley-Vroman, R. & Yoshinaga, N. (1992). Broad and narrow constraints on the english dative alternation: Some fundamental differences between native speakers and foreign language learners. *University of Hawai'i Working Papers in ESL*, 11, 157–199.

Bloch, B. (1947). English verb inflection. *Language*, 23(4), 399–418. https://doi.org/10.2307/410300

Bloom, L. (1970). *Language development: Form and function in emerging grammar*. Cambridge, MA: The MIT Press.

Bloom, P. (1990). Subjectless sentences in child language. *Linguistic Inquiry*, 21(4), 491–504.

Bloom, P. (2000). *How children learn the meanings of words*. MIT Press, Cambridge, MA.

Blum, L. & Blum, M. (1975). Toward a mathematical theory of inductive inference. *Information and control*, 28(2), 125–155. https://doi.org/10.1016/S0019-9958(75)90261-2

Borer, H. & Wexler, K. (1987). The maturation of syntax. In Roeper, T. and Williams, E., (Eds.), *Parameter setting*, 123–172. Berlin: Reidel. https://doi.org/10.1007/978-94-009-3727-7_6

Bowerman, M. (1982). Reorganizational process in lexical and syntactic development. In Wanner, E. & Gleitman, L. R., (Eds.), *Language acquisition: The state of the art*, (pp. 319–346). Cambridge: Cambridge University Press.

Bowerman, M. & Croft, W. (2008). The acquisition of the English causative alternation. In Bowerman, M. & Brown, P., (Eds.), *Crosslinguistic perspectives on argument structure: Implications for learnability* (pp. 279–307). Mahwah: Erlbaum.

Boyd, J. K. & Goldberg, A. E. (2011). Learning what not to say: The role of statistical preemption and categorization in a-adjective production. *Language*, 87(1), 55–83. https://doi.org/10.1353/lan.2011.0012

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training japanese listeners to identify english /r/and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, 61(5), 977–985. https://doi.org/10.3758/BF03206911

Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press. https://doi.org/10.4159/harvard.9780674732469

Bush, R. R. and Mosteller, F. (1951). A mathematical model for simple learning. *Psychological Review*, 68(3):313–323. https://doi.org/10.1037/h0054388

Bybee, J. L. (2006). *Frequency of Use and the Organization of Language*. Oxford University Press, Oxford.

Chomsky, N. (1955). *The logical structure of linguistic theory*. Ms., Harvard University and MIT. Revised version published by Plenum, New York, 1975.

Chomsky, N. (1958). [Review of Belevitch 1956]. *Language*, 34(1), 99–105. https://doi.org/10.2307/411281

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: The MIT Press.

Chomsky, N. (1968). *Language and mind*. New York: Harcourt, Brace and World.

Chomsky, N. (1981). *Lectures in government and binding*. Dordrecht: Foris.

Chomsky, N. (2001). *Beyond explanatory adequacy*. MIT Working Papers in Linguistics, Cambridge, MA.

Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry*, 36(1), 1–22. https://doi.org/10.1162/0024389052993655

Chomsky, N. & Halle, M. (1968). *The sound pattern of English*. Cambridge, MA: The MIT Press.

Clahsen, H. (1999). Lexical entries and rules of languge: A multidisciplinary study of German inflection. *Behavioral and Brain Sciences*, 22, 991–1069. https://doi.org/10.1017/S0140525X99002228

Clahsen, H. & Muysken, P. (1986). The availability of universal grammar to adult and child learners-a study of the acquisition of German word order. *Second Language Research*, 2(2), 93–109.

Clahsen, H., Rothweiler, M., Woest, A., & Marcus, G. (1992). Regular and irregular inflection in the acquisition of German noun plurals. *Cognition*, 45, 225–255. https://doi.org/10.1016/0010-0277(92)90018-D

Cochran, B. P., McDonald, J. L., & Parault, S. J. (1999). Too smart for their own good: The disadvantage of a superior processing capacity for adult language learners. *Journal of Memory and Language*, 41(1), 30–58. https://doi.org/10.1006/jmla.1999.2633

Cohen, W. W. (1995). Fast effective rule induction. In *Proceedings of the Twelfth International Conference on Machine Learning, Lake Tahoe, California*.

Conwell, E. & Demuth, K. (2007). Early syntactic productivity: Evidence from dative shift. *Cognition*, 103(2), 163–179. https://doi.org/10.1016/j.cognition.2006.03.003

Cook, V. & Newson, M. (2014). *Chomsky's universal grammar*. Oxford: Blackwell.

Crain, S., Koring, L., & Thornton, R. (2016). Language acquisition from a biolinguistic perspective. *Neuroscience & Biobehavioral Reviews*.

Crain, S. & Nakayama, M. (1987). Structure dependence in grammar formation. *Language*, 63(3), 522–543. https://doi.org/10.2307/415004

Culicover, P. W. (1999). *Syntactic nuts: Hard cases, syntactic theory, and language acquisition*. Oxford: Oxford University Press.

Cutler, A. & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, 2(3–4), 133–142. https://doi.org/10.1016/0885-2308(87)90004-0

Dąbrowska, E. (2001). Learning a morphological system without a default: The Polish genitive. *Journal of Child Language*, 28(3), 545–574.

Demuth, K. (1989). Maturation and the acquisition of the Sesotho passive. *Language*, 65(1), 56–80. https://doi.org/10.2307/414842

Demuth, K. (1996). The prosodic structure of early words. In Morgan, J. L. & Demuth, K., (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, (pp. 171–186). Hove: Psychology Press.

Diessel, H. (2013). Construction grammar and first language acquisition. In Hoffmann, T. & Trousdale, G., (Eds.), *The Oxford handbook of construction grammar*, pages 347–364. Oxford University Press, Oxford.

Dijkstra, T. (2005). Bilingual visual word recognition and lexical access. *Handbook of Bilingualism: Psycholinguistic approaches*, (pp. 179–201). Oxford: Blackwell.

Ellis, N. C. & Larsen-Freeman, D. (2009). Constructing a second language: Analyses and computational simulations of the emergence of linguistic constructions from usage. *Language Learning*, 59(s1), 90–125. https://doi.org/10.1111/j.1467-9922.2009.00537.x

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, 48(1), 71–99. https://doi.org/10.1016/0010-0277(93)90058-4

Epstein, S. D., Flynn, S., & Martohardjono, G. (1996). Second language acquisition: Theoretical and experimental issues in contemporary research. *Behavioral and Brain Sciences*, 19(4), 677–714. https://doi.org/10.1017/S0140525X00043521

Estes, W. K. (1950). Toward a statistical theory of learning. *Psychological Review*, 57(2), 94. https://doi.org/10.1037/h0058559

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B., & Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, pp. i–185. https://doi.org/10.2307/1166093

Fodor, J. D. (2001). Parameters and the periphery: Reflections on syntactic nuts. *Journal of Linguistics*, 37(2), 367–392.

Fodor, J. D. & Crain, S. (1987). Simplicity and generality of rules in language acquisition. In MacWhinney, B., (Ed.), *Mechanisms of language acquisition*, (pp. 35–63). Hillsdale: Lawrence Erlbaum Associates.

Friederici, A. D. (2017). *Language in Our Brain: The Origins of a Uniquely Human Capacity*. Cambridge, MA: The MIT Press.

Gerken, L. (1994). A metrical template account of children's weak syllable omissions from multisyllabic words. *Journal of Child Language*, 21(03), 565–584. https://doi.org/10.1017/S0305000900009466

Gibson, E. & Wexler, K. (1994). Triggers. *Linguistic Inquiry*, 25(3), 407–454.

Gold, E. M. (1967). Language identification in the limit. *Information and Control*, 10, 447–474. https://doi.org/10.1016/S0019-9958(67)91165-5

Goldberg, A. E. (1995). *Constructions*. Chicago: University of Chicago Press.

Goldin-Meadow, S. & Mylander, C. (1998). Spontaneous sign systems created by deaf children in two cultures. *Nature*, 391(6664), 279–281. https://doi.org/10.1038/34646

Goldin-Meadow, S. & Yang, C. (2017). Statistical evidence that a child can create a combinatorial linguistic system without external linguistic input: Implications for language evolution. *Neuroscience and Biobehavioral Reviews*, 81(Part B), 150–157. https://doi.org/10.1016/j.neubiorev.2016.12.016

Gordon, P. (1985). Evaluating the semantic categories hypothesis: The case of the count/mass distinction. *Cognition*, 20(3), 209–242. https://doi.org/10.1016/0010-0277(85)90009-5

Gordon, P. (1988). Count/mass category acquisition: distributional distinctions in children's speech. *Journal of Child Language*, 15(1), 109–128. https://doi.org/10.1017/S0305000900012083

Grimshaw, J. (1990). *Argument structure*. Cambridge, MA: The MIT Press.

Grinstead, J. (2000). Case, inflection and subject licensing in child catalan and spanish. *Journal of Child Language*, 27(1), 119–155. https://doi.org/10.1017/S0305000999004043

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. https://doi.org/10.2307/415332

Guasti, M. T. (2004). *Language acquisition: The growth of grammar*. Cambridge, MA: The MIT Press.

Hadley, P. A., Rispoli, M., & Holt, J. K. (2017). Input subject diversity accelerates the growth of tense and agreement: Indirect benefits from a parent-implemented intervention. *Journal of Speech, Language, and Hearing Research*, 60(9), 2619–2635. https://doi.org/10.1044/2017_JSLHR-L-17-0008

Hadley, P. A. & Walsh, K. M. (2014). Toy talk: Simple strategies to create richer grammatical input. *Language, Speech, and Hearing Services in Schools*, 45(3), 159–172. https://doi.org/10.1044/2014_LSHSS-13-0055

Halle, M. (1973). Prolegomena to a theory of word formation. *Linguistic Inquiry*, 4(1), 3–16.

Halle, M. & Vergnaud, J.-R. (1987). *An essay on stress*. Cambridge, MA: The MIT Press.

Harley, H. & Miyagawa, S. (2016). Ditransitives. *Oxford Research Encyclopedia of Linguistics*. Ed. Retrieved 21 Aug. 2018, from http://linguistics.oxfordre.com/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-186.

Hart, B. & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MS: Paul H Brookes Publishing.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298(5598), 1569–1579. https://doi.org/10.1126/science.298.5598.1569

Hayes, B. (1982). Extrametricality and English stress. *Linguistic Inquiry*, 13(2), 227–276.

Herrnstein, R. J. & Loveland, D. H. (1975). Maximizing and matching on concurrent ratio schedules. *Journal of the Experimental Analysis of Behavior*, 24, 107–116. https://doi.org/10.1901/jeab.1975.24-107

Hoff, E. (2014). *Language development*. Belmont: Wadsworth Cengage Learning.

Hornstein, N. & Lightfoot, D. (1981). Introduction. In Hornstein, N. & Lightfoot, D., (Eds.), *Explanation in linguistics: The logical problem of language acquisition*, (pp. 9–31). Oxford: Longman.

Hudson Kam, C. L. & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, 1(2), 151–195. https://doi.org/10.1080/15475441.2005.9684215

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, 27(2), 236. https://doi.org/10.1037/0012-1649.27.2.236

Hyams, N. (1986). *Language acquisition and the theory of parameters*. Dordrecht: Reidel. https://doi.org/10.1007/978-94-009-4638-5

Hyams, N. (1991). A reanalysis of null subjects in child language. In Weissenborn, J., Goodluck, H., & Roeper, T., (Eds.), *Theoretical issues in language acquisition: Continuity and change in development*, (pp. 249–268). Hove: Psychology Press.

Hyams, N. & Wexler, K. (1993). On the grammatical basis of null subjects in child language. *Linguistic Inquiry*, 24(3), 421–459.

Ibbotson, P. & Tomasello, M. (2016). Evidence rebuts Chomsky's theory of language learning. *Scientific American*, 315(5).

Inagaki, S. (1997). Japanese and Chinese learners' acquisition of the narrow-range rules for the dative alternation in English. *Language Learning*, 47(4), 637–669. https://doi.org/10.1111/0023-8333.00024

Ionin, T., Zubizarreta, M. L., & Bautista Maldonado, S. (2008). Sources of linguistic knowledge in the second language acquisition of english articles. *Lingua*, 118(4), 554–576. https://doi.org/10.1016/j.lingua.2006.11.012

Jackendoff, R. S. (1990). *Semantic structures*. Cambridge, MA: The MIT Press.

Jelinek, F. (1998). *Statistical methods for speech recognition*. Cambridge, MA: The MIT Press.

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, 21(1), 47–77. https://doi.org/10.1093/applin/21.1.47

Judy, T. (2011). L1/L2 parametric directionality matters: More on the null subject parameter in l2 acquisition. *EUROSLA Yearbook*, 11(1), 165–190. https://doi.org/10.1075/eurosla.11.10jud

Kanno, K. (1997). The acquisition of null and overt pronominals in japanese by english speakers. *Second Language Research*, 13(3), 265–287. https://doi.org/10.1191/026765897673070746

Kareev, Y. (1995). Through a narrow window: Working memory capacity and the detection of covariation. *Cognition*, 56(3), 263–269. https://doi.org/10.1016/0010-0277(95)92814-G
Kim, Y.-J. (2000). Subject/object drop in the acquisition of korean: A cross-linguistic comparison. *Journal of East Asian Linguistics*, 9(4), 325–351. https://doi.org/10.1023/A:1008304903779
Kowalski, A. & Yang, C. (2012). Verb islands in child and adult language. In Biller, A. K., Chung, E. Y., & Kimball, A. E., (Eds.), *BUCLD 36: Proceedings of the 36th annual Boston University Conference on Language Development*, (pp. 281–289), Somerset, MA: Cascadilla Press.
Krifka, M. (1999). Manner in dative alternation. In *West Coast Conference on Formal Linguistics*, 18, 260–271.
Kroch, A. (1989). Reflexes of grammar in patterns of language change. *Language Variation and Change*, 1(3), 199–244. https://doi.org/10.1017/S0954394500000168
Kučera, H. & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence: Brown University Press.
Labov, W. (1995). The case of the missing copula: The interpretation of zeros in African American English. In Gleitman, L. R. & Liberman, M., (eds.), *An invitation to cognitive science, Vol. 1: Language*, (pp. 25–54). Cambridge, MA: The MIT Press.
Lakoff, G. (1970). *Irregularity in syntax*. New York: Holt, Rinehart and Winston.
Legate, J. A. & Yang, C. (2007). Morphosyntactic learning and the development of tense. *Language Acquisition*, 14(3), 315–344. https://doi.org/10.1080/10489220701471081
Legate, J. A. & Yang, C. (2013). Assessing child and adult grammar. In Berwick, R. & Piattelli-Palmarini, M., (Eds.), *Rich languages from poor inputs: In honor of Carol Chomsky*, (pp. 168–182). Oxford: Oxford University Press.
Levin, B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press.
Lieven, E. V., Pine, J. M., & Barnes, H. D. (1992). Individual differences in early vocabulary development: Redefining the referential-expressive distinction. *Journal of Child Language*, 19(02), 287–310. https://doi.org/10.1017/S0305000900011429
Lignos, C. & Yang, C. (2016). Morphology and language acquisition. In Hippisley, A., & Stump, G.. (Eds.), *The Cambridge Handbook of Morphology*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781139814720.027
Lin, L. I. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, 45(1), 255–268. https://doi.org/10.2307/2532051
Liu, D. & Gleason, J. B. (2002). Acquisition of the article the by non-native speakers of English: An analysis of four nongeneric uses. *Studies in Second Language Acquisition*, 24, 1–26. https://doi.org/10.1017/S0272263102001018
MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. NJ: Lawrence Erlbaum, 3rd edition.
Mandelbrot, B. (1953). An informational theory of the statistical structure of language. In Jackson, B. W., (Ed.), *Communication theory*, Vol. 84, 486–502.
Marchman, V. A. & Bates, E. (1994). Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language*, 21(2), 339–366. https://doi.org/10.1017/S0305000900009302
Marcus, G., Pinker, S., Ullman, M. T., Hollander, M., Rosen, J., & Xu, F. (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development*. Chicago: University of Chicago Press.

Markson, L. & Bloom, P. (1997). Evidence against a dedicated system for word learning in children. *Nature*, 385(6619), 813–815. https://doi.org/10.1038/385813a0

Mazurkewich, I. & White, L. (1984). The acquisition of the dative alternation: Unlearning overgeneralizations. *Cognition*, 16(3), 261–283. https://doi.org/10.1016/0010-0277(84)90030-1

McClelland, J. L. & Patterson, K. (2002). Rules or connections in past-tense inflections: What does the evidence rule out? *Trends in Cognitive Sciences*, 6(11), 465–472. https://doi.org/10.1016/S1364-6613(02)01993-9

Miller, G. A. (1957). Some effects of intermittent silence. *American Journal of Psychology*, 70(2), 311–314. https://doi.org/10.2307/1419346

Miller, J. F. (1981). *Assessing language production in children: Experimental procedures*. London: Edward Arnold.

Miller, K. L. & Schmitt, C. (2012). Variable input and the acquisition of plural morphology. *Language Acquisition*, 19(3), 223–261.

Mitchell, T. M. (1982). Generalization as search. *Artificial Intelligence*, 18(2), 203–226. https://doi.org/10.1016/0004-3702(82)90040-6

Newport, E. (1990). Maturational constraints on language learning. *Cognitive Science*, 14(1), 11–28. https://doi.org/10.1207/s15516709cog1401_2

Newport, E. L. & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, 48(2), 127–162. https://doi.org/10.1016/S0010-0285(03)00128-2

Nida, E. A. (1949). *Morphology: the descriptive analysis of words*. Ann Arbor: University of Michigan Press, 2nd edition.

Osherson, D. N. & Smith, E. E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition*, 9(1), 35–58. https://doi.org/10.1016/0010-0277(81)90013-5

Perdue, C., (Ed.) (1993). *Adult language acquisition: Field methods*, Vol. 1. Cambridge: Cambridge University Press.

Pérez-Leroux, A. T. & Glass, W. R. (1999). Null anaphora in Spanish second language acquisition: Probabilistic versus generative approaches. *Second Language Research*, 15(2), 220–249. https://doi.org/10.1191/026765899676722648

Pesetsky, D. (1995). *Zero syntax: Experiencer and Cascade*. Cambridge, MA: MIT Press.

Phinney, M. (1987). The pro-drop parameter in second language acquisition. In Roeper, T. & Williams, E., (Eds.), *Parameter setting*, pp. 221–238. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-009-3727-7_10

Pierce, A. (1992). *Language acquisition and syntactic theory: A comparative analysis of French and English*. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-011-2574-1

Pine, J. M., Freudenthal, D., Krajewski, G., & Gobet, F. (2013). Do young children have adult-like syntactic categories? Zipf's law and the case of the determiner. *Cognition*, 127(3), 345–360. https://doi.org/10.1016/j.cognition.2013.02.006

Pine, J. M. & Lieven, E. V. (1997). Slot and frame patterns and the development of the determiner category. *Applied Psycholinguistics*, 18(2), 123–138. https://doi.org/10.1017/S0142716400009930

Pinker, S. (1979). Formal models of language learning. *Cognition*, 7(3), 217–283. https://doi.org/10.1016/0010-0277(79)90001-5

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press.

Pinker, S. (1995). Why the child holded the baby rabbit: A case study in language acquisition. In Gleitman, L. R. & Liberman, M., (Eds.), *An invitation to cognitive science, Vol. 1: Language*, pp. 107–133. Cambridge, MA: MIT Press.

Pinker, S. (1999). *Words and rules: The ingredients of language*. New York: Basic Books.

Pinker, S. & Ullman, M. T. (2002). The past and future of the past tense. *Trends in Cognitive Science*, 6(11), 456–463. https://doi.org/10.1016/S1364-6613(02)01990-3

Pizzuto, E. & Caselli, M. C. (1994). The acquisition of Italian verb morphology in a cross-linguistic perspective. In Levy, Y., (Ed.), *Other children, other languages*, pp. 137–187. Hillsdale, NJ: Lawrence Erlbaum.

Pullum, G. K. & Wilson, D. (1977). Autonomous syntax and the analysis of auxiliaries. *Language*, 53(4), 741–788. https://doi.org/10.2307/412911

Roberts, J. (1997). Acquisition of variable rules: A study of (-t, d) deletion in preschool children. *Journal of Child Language*, 24(2), 351–372. https://doi.org/10.1017/S0305000997003073

Rothman, J. & Slabakova, R. (2017). The generative approach to sla and its place in modern second language studies. *Studies in Second Language Acquisition*, 1–26. https://doi.org/10.1017/S0272263117000134

Saffran, J. R., Aslin, R. N., & Newport, E. (1996). Statistical learning by by 8-month-old infants. *Science*, 274, 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Sakas, W. G. & Fodor, J. D. (2012). Disambiguating syntactic triggers. *Language Acquisition*, 19(2), 83–143. https://doi.org/10.1080/10489223.2012.660553

Sakas, W. G., Yang, C., & Berwick, R. (2017). Parameter setting is feasible. *Linguistic Analysis*, 41(1).

Sapir, E. (1928). *Language: An introduction to the study of speech*. New York: Harcourt Brace.

Schuler, K. (2017). *The acquisition of productive rules in child and adult language learners*. PhD thesis, Georgetown University, Washington, D.C.

Schuler, K., Yang, C., & Newport, E. (2016). Testing the Tolerance Principle: Children form productive rules when it is more computationally efficient to do so. In *The 38th Cognitive Society Annual Meeting*, Philadelphia, PA.

Schwartz, B. D. & Sprouse, R. (1994). Word order and nominative case in nonnative language acquisition: a longitudinal study of (L1 turkish) german interlanguage. In Hoekstra, T. & Schwartz, B. D., (Eds.), *Language acquisition studies in generative grammar*, pp. 317–368. Amsterdam: John Benjamins. https://doi.org/10.1075/lald.8.14sch

Schwartz, B. D. & Sprouse, R. A. (1996). L2 cognitive states and the full transfer/full access model. *Second language research*, 12(1), 40–72. https://doi.org/10.1177/026765839601200103

Silvey, C. & Christodoulopoulos, C. (2016). Children's production of determiners as a test case for innate syntactic categories. In Roberts, S., Cuskley, C., McCrohon, L., Barcelô-Coblijn, L., Fehér, O., & Verhoef, T., (Eds.), *The Evolution of Language: Proceedings of the 11th International Conference (EVOLANGX11)*.

Slabakova, R. (2008). *Meaning in the second language*, Vol. 34. Walter de Gruyter. https://doi.org/10.1515/9783110211511

Smith, J., Durham, M., & Fortune, L. (2009). Universal and dialect-specific pathways of acquisition: Caregivers, children, and t/d deletion. *Language Variation and Change*, 21(1), 69–95. https://doi.org/10.1017/S0954394509000039

Smith, N. V. & Tsimpli, I.-M. (1995). *The mind of a savant: Language learning and modularity*. Oxford: Blackwell Publishing.

Snape, N. (2008). Resetting the Nominal Mapping Parameter: Definite article use and the count-mass distinction in L2 English. *Bilingualism: Language and Cognition*, 11(1), 63–79. https://doi.org/10.1017/S1366728907003215

Solomonoff, R. J. (1964). A formal theory of inductive inference. Part I. *Information and control*, 7(1), 1–22. https://doi.org/10.1016/S0019-9958(64)90223-2

Stefanowitsch, A. (2008). Negative entrenchment: A usage-based approach to negative evidence. *Cognitive Linguistics*, 19(3), 513–531. https://doi.org/10.1515/COGL.2008.020

Stevens, J., Trueswell, J., Yang, C., & Gleitman, L. (2016). The pursuit of word meanings. *Cognitive Science*, 41, 638–676. https://doi.org/10.1111/cogs.12416

Straus, K. J. (2008). *Validation of a probabilistic model of language acquisition in children*. PhD thesis, Northeastern University.

Suppes, P. (1974). The semantics of children's language. *American Psychologist*, 29(1), 103–114. https://doi.org/10.1037/h0036026

Terrace, H. S. (1987). *Nim: A chimpanzee who learned sign language*. Columbia University Press.

Terrace, H. S., Petitto, L.-A., Sanders, R. J., & Bever, T. G. (1979). Can an ape create a sentence? *Science*, 206(4421), 891–902. https://doi.org/10.1126/science.504995

Tettamanti, M., Alkadhi, H., Moro, A., Perani, D., Kollias, S., & Weniger, D. (2004). Neural correlates for the acquisition of natural language syntax. *NeuroImage*, 17, 700–709. https://doi.org/10.1006/nimg.2002.1201

Tokowicz, N. (2014). *Lexical processing and second language acquisition*. London: Routledge.

Tomasello, M. (1992). *First verbs: A case study of early grammatical development*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511527678

Tomasello, M. (2000a). Do young children have adult syntactic competence? *Cognition*, 74(3), 209–253. https://doi.org/10.1016/S0010-0277(99)00069-4

Tomasello, M. (2000b). First steps toward a usage-based theory of language acquisition. *Cognitive linguistics*, 11(1/2), 61–82.

Tomasello, M. (2003). *Constructing a language*. Cambridge, MA: Harvard University Press.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, 73(2), 89–134. https://doi.org/10.1016/S0010-0277(99)00032-3

Valian, V. (1991). Syntactic subjects in the early speech of American and Italian children. *Cognition*, 40(1), 21–81. https://doi.org/10.1016/0010-0277(91)90046-7

Valian, V., Solt, S., & Stewart, J. (2009). Abstract categories or limited-scope formulae? The case of children's determiners. *Journal of Child Language*, 36(4), 743–778. https://doi.org/10.1017/S0305000908009082

Van Assche, E., Duyck, W., & Hartsuiker, R. J. (2012). Bilingual word recognition in a sentence context. *Frontiers in psychology*, 3, 174.

Wang, Q., Lillo-Martin, D., Best, C. T., & Levitt, A. (1992). Null subject versus null object: Some evidence from the acquisition of Chinese and English. *Language Acquisition*, 2(3), 221–254. https://doi.org/10.1207/s15327817la0203_2

Wexler, K. (1998). Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive stage. *Lingua*, 106(1), 23–79. https://doi.org/10.1016/S0024-3841(98)00029-1

Wexler, K. & Culicover, P. (1980). *Formal principles of language acquisition*. Cambridge, MA: MIT Press.

White, L. (1985). Is there a "logical problem" of second language acquisition? *TESL Canada Journal*, 2(2), 29–42. https://doi.org/10.18806/tesl.v2i2.460

White, L. (1989). *Universal grammar and second language acquisition*. Amsterdam: John Benjamins. https://doi.org/10.1075/lald.1
White, L. (1990). The verb-movement parameter in second language acquisition. *Language Acquisition*, 1(4), 337–360. https://doi.org/10.1207/s15327817la0104_2
White, L. (2003). *Second language acquisition and universal grammar*. Cambridge University Press. https://doi.org/10.1017/CBO9780511815065
Wiese, R. (1996). *The phonology of German*. Oxford: Clarendon.
Xu, F. & Pinker, S. (1995). Weird past tense forms. *Journal of Child Language*, 22(3), 531–556. https://doi.org/10.1017/S0305000900009946
Yang, C. (2000). Internal and external forces in language change. *Language Variation and Change*, 12(3), 231–250. https://doi.org/10.1017/S0954394500123014
Yang, C. (2002). *Knowledge and learning in natural language*. Oxford: Oxford University Press.
Yang, C. (2004). Universal grammar, statistics or both? *Trends in Cognitive Sciences*, 8(10), 451–456. https://doi.org/10.1016/j.tics.2004.08.006
Yang, C. (2006). *The infinite gift: How children learn and unlearn the languages of the world*. New York: Scribner.
Yang, C. (2012). Computational models of syntactic acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3(2), 205–213.
Yang, C. (2013a). Ontogeny and phylogeny of language. *Proceedings of the National Academy of Sciences*, 110(16), 6324–6327 https://doi.org/10.1073/pnas.1216803110
Yang, C. (2013b). Who's afraid of George Kingsley Zipf? Or: Do children and chimps have language? *Significance*, 10(6), 29–34. https://doi.org/10.1111/j.1740-9713.2013.00708.x
Yang, C. (2015a). For and against frequencies. *Journal of Child Language*, 42(2), 287–293. https://doi.org/10.1017/S0305000914000683
Yang, C. (2015b). Negative knowledge from positive evidence. *Language*, 91(4), 938–953. https://doi.org/10.1353/lan.2015.0054
Yang, C. (2016). *The price of linguistic productivity: How children learn to break rules of language*. Cambridge, MA: MIT Press.
Yang, C. (2017). Rage against the machine: Evaluation metrics in the 21st century. *Language Acquisition*, 24(2), 100–125. https://doi.org/10.1080/10489223.2016.1274318
Yang, C., Crain, S., Berwick, R. C., Chomsky, N., & Bolhuis, J. J. (2017). The growth of language: Universal grammar, experience, and principles of computation. *Neuroscience and Biobehavioral Reviews*, 81(Part B), 103–119. https://doi.org/10.1016/j.neubiorev.2016.12.023
Yang, C., Ellman, A., & Legate, J. A. (2015). Input and its structural description. In Ott, D. & Gallego, A., (Eds.), *50th anniversary of Noam Chomsky's Aspects of the Theory of Syntax*. MITWPL.
Yang, C. & Montrul, S. (2017). Learning datives: The tolerance principle in monolingual and bilingual acquisition. *Second Language Research*, 33(1), 119–144. https://doi.org/10.1177/0267658316673686
Yip, K. & Sussman, G. J. (1997). Sparse representations for fast, one-shot learning. In *Proceedings of the National Conference on Artificial Intelligence*, pp. 521–527.
Yu, C. & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, 18(5), 414–420. https://doi.org/10.1111/j.1467-9280.2007.01915.x
Zipf, G. K. (1949). *Human behavior and the principle of least effort: An introduction to human ecology*. Cambridge, MA: Addison-Wesley.

Author's address

Charles Yang
Department of Linguistics and Computer Science
University of Pennsylvania
3401 Walnut Street 315C
Philadelphia, PA 19104
USA

charles.yang@ling.upenn.edu

COMMENTARIES

Less is More

On the Tolerance Principle as a manifestation of Maximize Minimal Means

Theresa Biberauer[1,2,3]
[1]University of Cambridge / [2]Stellenbosch University / [3]University of the Western Cape

The formalist perspective on language acquisition Yang outlines may be 'personal' (Yang, 2018, p. 665), but it offers a valuable illustration of the spirit guiding much 21st-century generative work. Central here is the drive to "reduce[] the explanatory burden traditionally placed on innate linguistic parameters" (Yang, 2018, p. 670) by pursuing a three- (1a) rather than two-factor (1b) model of grammar acquisition and structure:

(1) a. Universal Grammar (UG) + Primary Linguistic Data (PLD) + general cognitive principles → an adult (L1) grammar
b. UG + PLD → an L1 grammar

Where (1b) foregrounded a richly specified UG, (1a) emphasizes a pared-down UG, and the grammar-shaping role of non-language-specific cognitive mechanisms, including "principles of data analysis… used in language acquisition and other domains", and "principles of efficient computation" (Chomsky, 2005, p. 6). Yang's Tolerance Principle (TP) offers an explicit formal proposal as to how a potential third-factor principle – which rests on humans' well-established statistical sensitivity (Yang, 2004) – might interact with linguistic input to account for aspects of grammar acquisition and knowledge. As Yang convincingly demonstrates, the TP facilitates real insight into how "Less" can be "More" for child-acquirers. My purpose here is to endorse the "Less is More" spirit of Yang's model, and to suggest how the "periphery"-oriented TP may be integrated with a similarly "Less is More"-oriented three-factors approach to "core" grammar, which potentially also sheds light on the difference between L1- and L2-acquirers.

The TP is strongly lexically oriented: the tolerance calculation requires the identification of a domain encompassing a specific number of lexical items to

https://doi.org/10.1075/lab.18080.bib
Linguistic Approaches to Bilingualism 8:6 (2018), 707–711.
ISSN 1879-9264 | E-ISSN 1879-9272

which a hypothesized generalization could apply. For the calculation to be meaningful – for the TP to drive acquisition – there additionally need to be exceptions to the hypothesized rule. These requirements allow us to situate the TP in relation to a recently proposed three-factor model, Biberauer's (2017) so-called *Maximize Minimal Means (MMM)*-model.

The MMM-model postulates that child-acquirers aim to maximally exploit the knowledge at their disposal at all stages of the acquisition process, including the pre-lexical stage. MMM itself is conceived as a non-language-specific learning bias, which drives acquirers to construct their grammars incrementally on the basis of knowledge accessible to them at a given point. The idea is essentially that lesser access to input to begin with – less input which qualifies as *intake* – allows acquirers to focus their attention on a sub-component of the input, which they are then able to master and harness as the basis for access to more complex, previously inaccessible aspects of the input.

Much evidence suggests that L1-acquisition does indeed progress in this "Goldilocks" manner, with acquirers systematically attending to input that is neither too simple nor too complex, but "just right" – thus paralleling what we see in other domains (e.g., vision; Kidd, Piatandosi, & Aslin, 2012). Consider, for example, the research demonstrating *in utero* and very early post-birth sensitivity to prosody (Gervain & Werker, 2008 give an overview). Prosody delivers various 'edge'-oriented cues that allow acquirers to begin to "chunk" the input-strings in accordance with the grammar of their input-language(s) long before they have any lexical knowledge. Among other things, prosody alerts pre-lexical infants to basic head-directionality ("OV" = 'strong-weak' vs "VO" = 'weak-strong') and, during the first half-year, to the distinction between content and functional items, leaving 6-month-olds with a content-item preference. Thereafter, more fine-grained details become available, with, for example, the distribution of consonants and vowels within already-identified linguistic chunks facilitating the articulation of acquirers' knowledge of, respectively, vocabulary and inflectional morphology. Significant components of L1 knowledge are thus in place *before* the TP can regulate acquisition.

With some vocabulary in place, acquirers discover the key distinction between the truly arbitrary, memorization-requiring form-meaning mappings that characterize *content* items – classic Saussurean arbitrariness – and the "higher-level" arbitrariness that defines the recurring form and distribution of functional elements within a system – e.g., the obligatory inflection of Italian verbs, or the systematic fronting difference between lexical and auxiliary verbs in English questions. Biberauer (2017) proposes that such 'systematic departures from Saussurean arbitrariness' cue the postulation of grammar-defining formal ([F])-features alongside content-item-defining semantic and phonological features (Chomsky, 1995). Two points are crucial: firstly, [F]s define the formal make-up of functional categories, and, hence, parametric variation, or "core grammar". Secondly, the postulation of [F]s, in accordance with MMM-driven Feature Economy ('Postulate as few [F]s as possible to account for observed regularities') and Input Generalization ('Maximize the use of already-postulated [F]s'), allows the acquirer to optimize their knowledge of system-defining regularities in a manner parallelling that driving the TP-governed postulation of "periphery"-oriented rules: both reflect a response to the acquirer's 'search for [memorization-limiting – TB] productive generalizations' (Yang, 2018, p. 681), which is, at base, driven by MMM. Further, to the extent that [F]s have both grammar- (Narrow Syntax) internal and realizational (PF) consequences – triggering Agree and Move operations, and determining placement and morphological realization – we can begin to refine our understanding of how "core" [F]-mediated regularities and TP-sanctioned peripheral rules relate to one another, and what acquisitional consequences one might expect.

Significantly, the conclusion that peripheral rules are more "surfacey", with the TP merely regulating realizational options, seems too simple. Consider the "size"-based emergent (non-UG-given) parameter typology (2), and the associated 'learning pathway' (3):

(2) For a given value v_i of a parametrically variant feature [F]:
  a. *Macroparameters*: All heads of the relevant type, e.g., all probes/phase heads, share v_i . Examples: harmonic head-finality, radical pro-drop.
  b. *Mesoparameters*: All heads in a given natural class, e.g., [+V] or all C-/T-heads, share v_i . Examples: Chinese-style head-finality, Romance-style pro-drop, Verb-Second.
  c. *Microparameters*: A small, lexically definable sub-class of functional heads, e.g., (modal) auxiliaries, subject clitics, show v_i . Examples: partial pro-drop, English-style Verb-Second.
  d. *Nanoparameters*: One/more individual lexical items is/are specified for v_i. Examples: Bavarian-style pro-drop, English-style Conditional Inversion.

(3) Does [F]-based P(roperty) characterise L(anguage)?
- NO: **macroparameter**
- YES: All relevant heads?
  - YES: **macroparameter**
  - NO: A natural-class subset of heads?
    - YES: **mesoparameter**
    - NO: A further restricted natural-class subset of heads?
      - YES: **microparameter**
      - NO: Only lexically specified items? **nanoparameter**

Key here is the idea that macro-, meso- and microparameters are formulated over natural classes, namely, categories defined over more/less detailed [F]-specifications. L1-acquisition is largely assumed to proceed "top-down", with the acquirer's initially underspecified grammar progressively becoming more articulated in [F] terms as more of the input becomes accessible, and earlier acquisition stages, accordingly, featuring various underspecified "shadow" phenomena. More restricted/specialized natural classes are featurally more complex, meaning that we expect "smaller" [F]-regulated phenomena to be fully acquired later than "bigger" ones. Importantly, "bigger" class-based parameters do not seem to require acquirers to attend to TP-relevant lexical (ir)regularities.

Nanoparameters are different, however: they govern non-[F]-unified individual lexical items, are acquired bottom-up, independently of the general top-down process, and may therefore be acquired early or late, depending on the complexity of the [F]s encoded on the nanoparametrically specified items. Because nanoparameters don't target an entire natural class – consider (standard) English Conditional Inversion (CI), which affects only a subset of English's seven past-marked auxiliaries: *had, were, should* – and because the class is quantifiable, they seem to exhibit the same profile as TP-sanctioned "peripheral" rules. And English CI in fact satisfies the TP ($N/lnN = 1.946$). In principle, given the definition of nanoparameter, this need not always be the case, however, leading us – seemingly correctly – to expect TP-regulated instability in the nanoparametric corners of grammar. The TP may thus regulate not just the productivity of regular rules like those discussed in Yang's work, but also the stability of certain types of exceptions, the nanoparameters, which often instantiate "left-overs" from a more productive system.

What does this imply for our understanding of the differences between L1- and (adult) L2-acquisition? Yang suggests that adult learners' larger vocabularies effectively undermine their ability to harness the TP with an L1-acquirer's effectiveness, leaving them unable to reap "Less is More"-type benefits. From an MMM perspective, this may be part of the problem; more generally, however, it seems that adults' abundance of highly-developed resources – language-specific and general-cognitive – compromises their ability to "get back to basics" so as to build up their grammars incrementally, via the "Goldilocks"-route taken by L1-acquirers. If our discussion is on the right track, it will not just be the lexically-based and thus potentially TP-regulated aspects of grammar that L2-acquirers will achieve variable success with; invariant and thus non-TP-regulated "bigger" properties will also be undercut by the knowledge and biases that adults bring to the task. And, if that is correct, three-factors-inspired formal work has much to offer to deepen our understanding of the surprising properties of L1-acquisition, mono- and bilingual, and of the nature of the challenge facing L2-learners.

References

Biberauer, T. (2017). Factors 2 and 3: a principled approach. *Cambridge Occasional Papers in Linguistics* 10: 38–65.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry* 36, 1–22. https://doi.org/10.1162/0024389052993655

Gervain, J. & Werker, J. (2008). How infant speech perception contributes to language acquisition. *Language and Linguistics Compass* 2(6), 1149–1170. https://doi.org/10.1111/j.1749-818X.2008.00089.x

Kidd, C., Piatandosi, S. & Aslin, R. (2012). The Goldilocks Effect: human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS ONE* 7(5), e36399. https://doi.org/10.1371/journal.pone.0036399

Yang, C. (2004). Universal Grammar, statistics, or both? *Trends in Cognitive Science* 10, 451–456. https://doi.org/10.1016/j.tics.2004.08.006

Yang, C. (2018). A formalist perspective on language acquisition. *Linguistic Approaches to Bilingualism* 6(6), 665–706.

Address for correspondence

Theresa Biberauer
University of Cambridge
Computer Science and Technology Department, The Computer Laboratory
William Gates Building
15 JJ Thompson Avenue, Cambridge, CB3 0FD
United Kingdom

mtb23@cam.ac.uk
https://orcid.org/0000-0003-3840-7618

Publication history

Date received: 20 August 2018
Date accepted: 21 August 2018

Evaluating Yang's algorithms

An outline

Cécile De Cat
University of Leeds

Yang (2018) proposes three algorithms to inform language acquisition research: one to predict within-learner grammar competition in light of variability in the input (the Variational Model), one to assess combinatorial productivity in light of the distribution of the lexical items involved (the extension to Zipf's law), and one to quantify the minimum amount of positive evidence that can support the postulation of a rule in light of the number of lexical items in the learner's vocabulary that the rule applies to (the Tolerance Principle, and its corollary, the Sufficiency Principle).

It has become uncontroversial that the computation of frequencies informs the language acquisition process, both in first and in second language acquisition. The acknowledgement of the role of the Third Factor (Chomsky, 2005) has rightly sparked a reappraisal of the learner-internal determinants of language acquisition. Yang contends that his algorithms can help inform the debate between what is innate and what is emergent, and provide answers to the learnability question (see Lidz & Gagliardi, 2015, for an alternative model that takes sensitivity to statistical-distributional features of the input into account and proposes an inference mechanism that links the representations provided by UG with the data of experience).

In this commentary, I contribute a number of questions to inform the appraisal of Yang's proposal. I focus on the assumptions that underlie the proposed algorithms, the domain of applicability of these algorithms, the need for embedding in a coherent theory of language acquisition, and predictions regarding bilingual language acquisition.

1. Assumptions

The extension to Zipf's law and the Tolerance Principle are offered to "develop a principled quantitative interpretation of language" and provide a "statistically

https://doi.org/10.1075/lab.18066.de
Linguistic Approaches to Bilingualism 8:6 (2018), **712–716**.
ISSN 1879-9264 | E-ISSN 1879-9272

rigorous assessment of what constitutes evidence for a productive grammar" (p. 675), respectively. These are laudable aims. However, I would have expected from a keynote paper that the "principled" aspects be presented, at least in general terms. Take the extension to Zipf's law: if the probability of the combination of two classes of elements approximates the multiplication of their Zipfian distribution, how does this *demonstrate* that this distribution is not lexically-specific? What is the null hypothesis, and how can it be assessed? In the case of the Tolerance Principle, we are referred to another paper for empirical motivation, and asked to accept without any explanation that the maximal number of exceptions to a rule should equate to the number of items (in the candidate set) divided by the natural log of that number of items. The natural log has been used elsewhere to model growth (e.g., in finance), but how exactly does this address the problem at hand? Any why is the estimation based on the *learner's* vocabulary size, rather than frequencies in the input?

A second point in need of clarification is that of error margins. To what extent does Zipf's law accurately predict the distribution of lexical items in corpora of adult-child interactions (Lestrade, 2017; Ryland Williams et al., 2015)? How accurately does the multiplication of the Zipfian probability of two elements predict their actual distribution in a data set? How can we tell with confidence that the predicted and the observed values are "statistically indistinguishable", as claimed with respect to the data presented in Table 1 in the keynote paper? We are promised rigorous methods, but their application feels rather sketchy, at least as presented in the keynote paper.

2. Applicability and implications for a theory of language acquisition

The extension to Zipf's law is argued to assess syntactic productivity, which, in turn, Yang interprets as an indication that the child has learned a "particular rule of the English grammar" (p. 678). However, demonstrating the combinatorial productivity of determiner + noun combinations is not the same as demonstrating that the child has acquired the constraints that condition determiner use in the target language (as Yang does concede). At best, what the algorithm might be able to assess is whether the child "knows" that merge can target D + N (which assumes the existence of these categories in the child's grammar).

The principles of Tolerance and Sufficiency are stated in relation to the number of relevant items in the learner's vocabulary. Their applicability is illustrated with respect to rules that are lexically conditioned: past tense formation (where a morphological rule applies to a verb stem in most but not all of the paradigm) and dative alternation (which is partly conditioned by the lexical-semantic properties of the verb). Is the Tolerance Principle intended to apply to rules that are not so lexically conditioned? How should one determine the lexical class of elements relevant to model the learnability of e.g., wh-movement, scrambling, or Principle B?

In both cases, the actual remit of the algorithm is lexically defined. To what extent does this require the language "rules" they target to be formulated in a lexicalist framework too? At present, it seems to me that the types of phenomena they might apply to are quite limited. This puts a damper on the claims that the Tolerance Principle can explain "how children form the rules of language even in the face of exceptions" (p. 666).

Also unclear is how the algorithms proposed in the second part of the paper are supposed to interact with the Variational Model. When the learner's vocabulary includes sufficient evidence in the input to support the postulation of a rule, is the rule then assigned a probability value based on the robustness of the cues in the input?

One of the attractions of the Variational Model is that it exploits grammatical rules that are *independently* defined by grammatical theory (e.g., parameters). These define the hypothesis space that informs learner choices. As Yang points out, the theory of parameters is currently under review. But the status of what constitutes the child's hypothesis space is not a mere issue of terminology, as he suggests. And it is quite unclear at this point how the Tolerance Principle can "reduce the explanatory burden traditionally placed on innate linguistic patterns" (p. 670), if there is no clear demarcation of parameters. I see more promise in the approaches of Biberauer (2017), Biberauer and Roberts (2017), Kazakov et al. (2018), Biberauer (this volume).

3. Predictions for bilingualism

Yang contends that "rule learning is easier, and more tolerant of exceptions, when the learner has a smaller set of items in their vocabulary" (p. 692), and he argues this is one of the key reasons why children are better language learners than adults. The implication is that the optimal learner is a child with a small vocabulary. Bilingual children provide an ideal test case for this claim: their vocabulary in each language tends to be more limited than that of their age-matched monolingual peers. Yang's proposal therefore predicts that bilingual children with small vocabularies should be more efficient at "rule learning". Assuming that rule learning requires efficient processing of the input, existing evidence suggests the opposite might be true: Marchman, Fernald and Hurtado (2010) show that, in 2;6 year-old bilinguals, the efficiency of online processing in one language is significantly related to vocabulary size in that language, even after controlling for processing speed and vocabulary size in the other language.

4. Final remarks

It is to Yang's credit to propose a new approach to the issue of how little evidence is enough to support language acquisition. Ultimately, what we need to understand is how children acquire knowledge that is richer than the input they experience, and how much learning is inferred rather than learned. As Yang points out, UG alone does not explain this: it has to be combined with a language acquisition device. Yang invites us to entertain precise hypotheses regarding the role of distributional frequencies in language acquisition. This, I believe, could bring us closer to the identification of objective thresholds for genuine Poverty of the Stimulus phenomena (which cannot be learned from the input), and from there a better understanding of the role of UG in language acquisition.

References

Biberauer, T. (2017). *Going beyond the input: Three factors and language acquisition, variation and change*. (Talk given at the University of Leeds, October 2017).

Biberauer, T., & Roberts, I. (2017). Parameter setting. In A. Ledgeway & I. Roberts (Eds.), *The cambridge handbook of historical syntax* (pp. 134–162). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781107279070.008

Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry*, 36(1), 1–22. https://doi.org/10.1162/0024389052993655

Kazakov, D. L., Cordoni, G., Algahtani, E., Ceolin, A., Irimia, M.-A., Kim, S.-S., … Longobardi, G. (2018). Learning implicational models of universal grammar parameters. In C. Cuskley, M. Flaherty, H. Little, L. McCrohon, A. Ravignani, & T. Verhoef (Eds.), *The evolution of language: Proceedings of the 12th international conference (evolangxii)*. Torun, Poland: Online at http://evolang.org/torun/proceedings/papertemplate.html?p=176 https://doi.org/10.12775/3991-1.048

Lestrade, S. (2017). Unzipping zipf's law. *Plos One*, 12(8), e0181987. https://doi.org/10.1371/journal.pone.0181987

Lidz, J., & Gagliardi, A. (2015). How nature meets nurture: Universal grammar and statistical learning. *Annual Review of Linguistics*, 1(1), 333–353. https://doi.org/10.1146/annurev-linguist-030514-125236

Marchman, V. A., Fernald, A., & Hurtado, N. (2010). How vocabulary size in two languages relates to efficiency in spoken word recognition by young spanish-english bilinguals. *Journal of Child Language*, 37(4), 817–840. https://doi.org/10.1017/S0305000909990055

Ryland Williams, J., Lessard, P. R., Desu, S., Clark, E. M., Bagrow, J. P., Danforth, C. M., & Sheridan Dodds, P. (2015). Zipf's law holds for phrases, not words. *Scientific Reports*, 5, 12209. https://doi.org/10.1038/srep12209

Yang, C. (2018). A formalist perspective on language acquisition. *Linguistic Approaches to Bilingualism*, 8(6), 665–706.

Address for correspondence

Cécile De Cat
University of Leeds
Linguistics & Phonetics
Woodhouse Lane
Leeds, LS2 9JT
United Kingdom

c.decat@leeds.ac.uk
https://orcid.org/0000-0003-0044-0527

Publication history

Date received: 11 July 2018
Date accepted: 21 August 2018

Input and the acquisition of productive grammatical knowledge

Vocabulary size as missing link?

Christine Dimroth
University of Münster

Recent emergentist approaches to second language acquisition like De Bot, Lowie and Verspoor (2007) attempt to tackle a multitude of factors that are known to have an impact on aspects of second language acquisition but are very different in nature, e.g., the structure of L1 and L2, input frequencies, the learning situation, individual learner differences, and many more. Against this background, Charles Yang's Variational Learning Model comes across as refreshingly focused. Yang's model aims for easily countable ingredients and falsifiable predictions about the relation between input, vocabulary size and the acquisition of productive grammatical knowledge, i.e., a core issue in first and second language acquisition research (see Hulstijn, 2015). By claiming that child language is productive from early on and acquisition is mainly driven by statistical learning, Yang combines credos typically associated with generative and usage-based theories respectively. His approach is in many ways different from the one I find most interesting and fruitful, in particular because acquisition is described as the task to discover and gauge the productivity of isolated formal patterns as if they had no function in either the target language, or the developing grammar.

Instead of laying out how I would go about investigating these questions or blaming Yang for not sharing my interests, in the following I will try to stick with Yang's overall perspective. First, I briefly mention three issues from his article that I find important, but where I find Yang's reasoning difficult to understand, namely, the role of a theory, evidence from the input, and the notion of 'command'. In the second part, I will home in on Yang's proposal that the main difference between child L1 and adult L2 learners is the comparably rapid vocabulary growth and the (partly detrimental) consequences for the statistical learning of grammatical properties in the latter group.

https://doi.org/10.1075/lab.18057.dim
Linguistic Approaches to Bilingualism 8:6 (2018), 717–721.
ISSN 1879-9264 | E-ISSN 1879-9272

The role of a theory: When Yang (2018) states that "methodological rigor should override the researcher's preference for the mode of explanation," (p. 665–666) he suggests a primarily data-driven approach. Some of Yang's assumptions, e.g., that native adult grammars are characterized by general linguistic rules, are, however, not theoretically innocent at all. I have no principled problem with this particular way of putting things, but it should be clear that the hypotheses tested with the help of mathematical models are (and should be!) derived from theoretical considerations.

Evidence from the input: Given the idea that learners can entertain target-deviant hypotheses at the outset, Yang's Variational Learning model derives its predictive power from the identification and quantification of disambiguating evidence available in the learners' input. In the case of early subjectless utterances in child English, exposure to highly frequent overt subjects should in principle drive learning in the same way that the exposure to highly frequent regular past tense forms does. Yang states that "it is easy to see that if such rules were used, English learning children should quickly converge to the target grammar" (p. 689) but – given the observed slow acquisition of obligatory subjects – concludes that infrequent expletives must instead be the relevant evidence. Yet this reasoning seems somewhat circular, unless one assumes right from the start that overt thematic subjects cannot distinguish a topic-drop-grammar from an obligatory-subject-grammar, but this is again a theory-driven statement.

The notion of 'command': Referring to L2-literature, Yang claims that "the acquisition of pro drop and topic drop appears "easier" than the acquisition of the obligatory subject use" and that "even beginner L1 English learners of Spanish show excellent command of pro drop" (p. 689). However, L2 learners' excellent command of pro drop is unlikely to include the syntactic and pragmatic knowledge guiding *the choice* between zero and overt pronouns in pro drop languages. Based on an array of empirical studies on the acquisition of form-function associations in pro drop grammars, Sorace concludes: "The common finding is that bilinguals tend to produce and accept overt subject pronouns referring to a pragmatically inappropriate topic subject antecedent significantly more often than monolinguals" (2011, 4). Broad notions like 'command' seem to underdetermine the nature of the knowledge learners have and are able to use in production and comprehension.

In the remainder of this comment I will focus on Yang's application of the Variational Learning model, and, in particular, the Tolerance Principle to L2 acquisition. This principle asserts that "all else being equal" – and this is precisely what Yang assumes – "productive rules are *easier* to detect for learners who have access to *less* input data" (p. 682). By input data, Yang actually refers to vocabulary size due to slower or faster lexical acquisition. This is reminiscent of Newport's (1990) *less-is-more* hypothesis ascribing non-native outcomes of L2 acquisition to older learners' counterproductive increase in memory and information processing capacities. The following two questions concerning this approach will be addressed in turn: How do learning and unlearning play out in a developmental perspective, and why is prior L1 knowledge considered irrelevant for L2 acquisition?

Yang's general idea is that the proportions of regular and productive features vs. exceptions change as a function of the sample size available to the learners. Growing sample sizes due to development affect acquisition targets in variable ways: The productivity of the *-ed* suffix for regular past tense marking cannot be detected in a small sample; the rule only becomes productive when more vocabulary is acquired. The double object construction is wrongly considered fully productive when calculated on the basis of a small vocabulary, but productivity will be restricted when proportions change due to lexical growth. The productivity of article-noun combinations can (only) be detected in a small sample and is subsequently unaffected by the availability of a larger vocabulary. It remains unclear, however, why (for example) the over-productivity of double object constructions can be unlearned when the relevant proportions change in a larger sample, but the productivity of article-noun combinations that adults are presumably unable to detect in a too-large sample remains unchanged when child vocabularies grow.

With the help of the Tolerance Principle, Yang explains how child learners of L1 English manage to discover the productivity of determiner-noun-combinations even though only around 30% of the nouns encountered in the input are attested with both definite and indefinite articles. This type of learning problem arises under the assumption that learners rely on the form of the input NPs without paying attention to the contexts in which definite vs. indefinite articles occur (but see Serratrice & Allen, 2015). This is hardly plausible for L2 learners who can transfer knowledge about the function and the productivity of (in)definiteness-markers onto the new language if their L1 is sufficiently similar. The two groups of L2 learners of English from the ESF corpus analyzed by Yang have Italian (an article system similar to English) and Punjabi (no article system; Gill & Gleason, 1963) as their first languages. Given that these learners face a different abductive learning task, their data should not be lumped together.

I will close with a more general remark concerning a core issue of Yang's paper to which I can fully subscribe: methodology matters, as different methodological choices can result in diametrically opposed answers given to very similar questions. Much like Yang, Dimroth and Haberzettl (2012) have investigated the growing productivity of grammatical knowledge in corpus data from different learner populations. Their study compares the development of German verb inflection in L1 learners and 7–9 year-old L2 learners with L1 Russian. Like the L2 adults investigated by Yang, the latter group can be assumed to dispose of advanced word learning resources and thus bigger sample sizes. Contrary to Yang, the learners' growing productivity of grammatical knowledge was measured longitudinally on the basis of morphological contrasts in unfolding mini-paradigms. Based on this conceptualization and measurement of productivity, Dimroth and Haberzettl found that the L2 learners outperformed the L1 learners with respect to the time needed for productive use of verbal morphology. Consequently, they concluded that *more (!) is more* in child L2 acquisition.

Which measures of productivity in learner corpora and input (Dimroth, 2018) one finds adequate and informative depends on one's theoretical assumptions. However, as long as each theoretical approach sticks to its own methods, not much progress can be made. In the long run, researchers will probably have to keep questions *and* data constant, compare different methods of analysis, and argue about the interpretation of the results in a collaborative effort.

References

de Bot, K., Lowie, W. M., & Verspoor, M. H. (2007). A Dynamic Systems Theory approach to second language acquisition. *Bilingualism: Language and Cognition*, 10, 7–21. https://doi.org/10.1017/S1366728906002732

Dimroth, C. (2018). Beyond statistical learning: Communication principles and language internal factors shape grammar in child and adult beginners learning Polish through controlled exposure. *Language Learning*. https://doi.org/10.1111/lang.12294

Dimroth, C., & Haberzettl, S. (2012). The Older the Better, or More is More: Language Acquisition in Childhood. In M. Watorek, S. Benazzo & M. Hickmann (Eds.), *Comparative Perspectives to Language Acquisition: A tribute to Clive Perdue* (pp. 324–350). Clevedon: Multilingual Matters.

Gill, H. S. & Gleason, H. A. (1963). *A Reference Grammar of Panjabi*. Hartford: Hartford Seminary Foundation.

Hulstijn, J. H. (2015). How Different Can Perspectives on L2 Development Be? *Language Learning*, 65, 210–232. https://doi.org/10.1111/lang.12096

Newport, E. (1990). Maturational constraints on language learning. *Cognitive Science*, 14, 11–28. https://doi.org/10.1207/s15516709cog1401_2

Serratrice, L. & Allen, S. (2015), (Eds.). *The Acquisition of Reference*. Amsterdam: John Benjamins. https://doi.org/10.1075/tilar.15

Sorace, A. (2011). Pinning down the concept of "interface" in bilingualism. *Linguistic Approaches to Bilingualism*, 1, 1–33. https://doi.org/10.1075/lab.1.1.01sor

Yang, C. (2018). A Formalist Perspective on Language Acquisition. *Linguistic Approaches to Bilingualism*, 8(6), 665–706.

Address for correspondence

Christine Dimroth
Germanistisches Institut
Westfälische Wilhelms-Universität Münster
Schlossplatz 34
48143 Münster
Germany

christine.dimroth@uni-muenster.de
https://orcid.org/0000-0003-0483-7839

Publication history

Date received: 24 June 2018
Date accepted: 15 August 2018

What is the role of L1 representations in a grammar-input model of L2 acquisition?

Laura Domínguez and Jorge González Alonso
University of Southampton / UiT The Arctic University of Norway

Yang's (2018) variational language acquisition model provides a promising framework for understanding how languages are acquired. The focus of Yang's epistemological paper is to extend his model to the domain of adult second language (L2) acquisition. His main proposal is that the interaction of target language input and universal, domain-specific mechanisms of rule-inferencing – summarised in his three equations – is sufficient to explain language acquisition. In this model, language acquisition mainly involves a process by which learners have to determine a set of productive rules from the available input. Yang acknowledges that the rule-inferencing mechanism that he proposes for child language acquisition does not fully explain why, unlike children, L2 learners are often unsuccessful. To address this issue, he proposes that learners' vocabulary size, which is bigger than children's, interferes with the inductive process. Having a large vocabulary raises the threshold for productivity, which, in turn, makes learning rules more challenging. Adult learners are, thus, disadvantaged in the acquisition process with respect to children, as adults are cognitively mature and know too much about language.

While we are sympathetic to Yang's mechanistic approach, in that it operates on strong hypotheses and thus consistently generates empirically falsifiable predictions, we believe that his take on adult non-native acquisition data somewhat disregards or downplays the fundamental role of L1 representations which already exist in the grammars of learners. This is despite the fact that decades of work, particularly since the 1990s, have established the pervasiveness of L1 transfer effects and the constraints they impose on L2 learnability – e.g., Full Transfer/Full Access (Schwartz & Sprouse, 1996); representational deficit hypotheses (Hawkins and Chan, 1997; Hawkins & Liszka, 2003; Tsimpli & Dimitrakopoulou, 2007); Feature Reassembly accounts (Lardiere, 2009; Choi & Lardiere, 2006). We would like to highlight that some effects of L1 representations in L2 acquisition are not incompatible with Yang's general approach. However, we believe that developing

https://doi.org/10.1075/lab.18060.dom
Linguistic Approaches to Bilingualism 8:6 (2018), 722–726.
ISSN 1879-9264 | E-ISSN 1879-9272

an extension of the model to non-native language acquisition without factoring in previous linguistic experience as a main variable is missing a large part of the picture that research into L2 acquisition has uncovered. In this commentary, we elaborate on the reasons to do so and provide evidence from some well-known L2 data that cannot be accounted for by Yang's model as it currently stands.

The three equations Yang proposes in the first part of the article efficiently conspire to derive sophisticated rules of, for example, the regular past tense in English. For Yang, the bulk of the task of constructing past tense during language acquisition falls on identifying the rules that allow verbs in English to be marked with '-ed'. What remains unclear, however, is how such rules are integrated into a broader grammatical system, above and beyond their immediate application in parsing; in our view, the model does not provide an explanation for how learners map the rules they learn onto the abstract features and categories that underlie syntactic operations. For instance, in inducing the rules of verbal inflection as it relates to the past tense, do learners acquire a [+tense] feature? Assuming that they do, how does this impact their representation of a Tense Phrase? And, crucially in our view, how is the mapping between these abstract categories and the language-specific functional morphology established?

These issues have non-trivial implications for the way in which child grammars develop into their adult state and, consequently – and importantly here – for the kind of initial conditions one implicitly or explicitly assumes in sequential adult bilingualism. Adult learners have already acquired at least one native grammar and so they already have access to language-specific rules and syntax-to-form mappings of abstract features – e.g., that number agreement can be established between nouns and adjectives, articles, demonstratives, verbs and one or more of their arguments, etc. The learning task in this case is, thus, subtly different to that of L1 acquisition even if all underlying mechanisms are one and the same: not all L2 abstract features need to be mapped anew onto the morphology; rather, for some of them, the L2 learner must determine how those very same mappings (beyond their morphophonological exponents) differ between the L1(s) and the L2. In other words, learners must (i) learn new rules for properties not instantiated in the L1(s); (ii) learn new rules for contexts in which the native grammar did not instantiate a given property/feature that is nevertheless present in other domains; and (iii) reconfigure the mappings of different features. Crucially, (ii) and (iii) are absent from the L1 learning task, because they stem from the availability of one or more previously acquired grammars. In the two scenarios which are unique to L2 learners, potential influence from the L1 has to be considered.

This is, indeed, what existing evidence from L2 acquisition literature has revealed over the last few decades, in particular, evidence from studies focusing on the acquisition of past tense. While speakers of languages which arguably lack morphosyntactic expressions of Tense (such as Chinese) often omit '-ed' in English, (Lardiere, 2006, 2009), Hawkins and Liszka (2003) report that German and Japanese learners of English show consistent targetlike use of past tense markings (German, Japanese and English being languages with similar morphosyntactic properties which relate to the Tense Phrase). Lardiere's longitudinal study of a Chinese (Mandarin and Hokkien) experienced learner of English in the US, known as Patty, similarly shows low use of past tense marking as Patty only marks past tense with '-ed' 34 per cent of the time, even though she has had ample opportunity to infer the underlying rule. According to Yang's assumptions, the rule responsible for marking past tense verbs with '-ed' is not productive in Patty's case. We see the logic in reaching this conclusion if one assumes that language learning is a process based on detection of productive rules from the distributional properties of the input. However, one must wonder why it is the case that speakers of languages without morphological expressions of tense have significantly greater difficulties in discovering the productive rule of past tense marking in English in the studies cited above (see Cabrelli Amaro et. al., 2017, for a different view). A further unresolved issue is that, whereas Patty does not always mark past tense (i.e., she has not figured out the corresponding productive rule in Yang's model), she shows higher use of other rule-based structures such as plural '-s' (58 per cent) and definite articles (84 per cent). Based on this evidence, two key remaining questions are, first, why the same individual internalized vocabulary can only determine a subset of the productive rules successfully (i.e., why the *add '-ed'* rule should be more challenging than the *add '-s'* rule), and, second, why Patty applies some of the learned rules optionally (only 58 per cent of plural '-s' are supplied), a phenomenon which is well documented in L2 acquisition (see also Slabakova, this volume).

We agree with Yang that rule learning on the basis of the available input is a fundamental part of the language acquisition process across contexts, but how those rules are mapped onto actual grammatical representations is crucial as well, and so is the (well-known) fact that L2 speakers are influenced by the grammars of their first language. In our view, assuming that one single mechanism of language learning (namely rule induction) can account for the whole of the L2/bilingual acquisition process does not do justice to the complexity of the phenomenon. L2 learners, unlike monolingual children, need to overcome the challenges that their own native language poses to the L2 learning task. Yang's model will need to accommodate the possibility of L1 influence in order to successfully account for (child and adult) bilingual acquisition.

References

Cabrelli Amaro, J., Campos-Dintrans, G., & Rothman, J. (2017). The Role of L1 Phonology in L2 Morphological Production: L2 English Past Tense Agreement from Spanish, Chinese and Japanese Natives. *Studies in Second Language Acquisition*, 1–25. 40(3), 503–527. https://doi.org/10.1017/S0272263117000122

Choi, M. H. & Lardiere, D. (2006). The interpretation of wh-in-situ in Korean second language acquisition. In Belletti, A., E. Bennati, C. Chesi, E. DiDomenico, and I. Ferrari (eds.), *Language Acquisition and Development: Proceedings of GALA 2005* (pp. 125–135). Newcastle, UK: Cambridge Scholars Publishing.

Hawkins, R., & Chan, C. Y. H. (1997). The partial availability of Universal Grammar in second language acquisition: The 'failed functional features hypothesis'. *Second Language Research*, 13(3), 187–226. https://doi.org/10.1191/026765897671476153

Hawkins, R., & S. Liszka. (2003). Locating the source of defective past tense marking in advanced L2 English speakers. In van Hout, R., A. Hulk, F. Kuiken and R. Towell (eds.), *The lexicon-syntax interface in second language acquisition* (pp. 1–44). Amsterdam: John Benjamins. https://doi.org/10.1075/lald.30.03haw

Lardiere, D. (2006). *Ultimate attainment in second language acquisition: A case study*. Mahwah, NJ: Laurence Erlbaum.

Lardiere, D. (2009). Some thoughts on the contrastive analysis of features in second language acquisition. *Second Language Research*, 25(2), 173–227. https://doi.org/10.1177/0267658308100283

Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the full transfer/full access model. *Second Language Research*, 12(1), 40–72. https://doi.org/10.1177/026765839601200103

Slabakova, R. (this volume). Back to our roots. *Linguistic Approaches to Bilingualism*, 8(6), 776–780.

Tsimpli, I. M., & Dimitrakopoulou, M. (2007). The interpretability hypothesis: Evidence from wh-interrogatives in second language acquisition. *Second Language Research*, 23(2), 215–242. https://doi.org/10.1177/0267658307076546

Yang, C. (2018). A formalist perspective on language acquisition. *Linguistic Approaches to Bilingualism*, 8(6), 665–706.

Address for correspondence

Laura Domínguez
University of Southampton
Modern Languages and Linguistics
Southampton SO17 1BJ
United Kingdom

L.Dominguez@soton.ac.uk

Co-author details

Jorge González Alonso
UiT The Arctic University of Norway
Department of Language and Culture
Tromsø 9019
Norway

jorge.gonzalez.alonso@uit.no

Publication history

Date received: 30 June 2018
Date accepted: 15 August 2018

The sufficiency principle hyperinflates the price of productivity

Adele E. Goldberg
Princeton University

Yang's (2016) *Price of Linguistic Productivity* (*PoLP*) and the target article in this volume (Yang, 2018) offer a proposed solution to the issue of how learners recognize when constructions (or "rules") may be used productively. For example, the English prefix, *pre-*, is productive insofar as it can be applied to any noun or proper name that can be construed to evoke a temporal onset (*pre-children, pre-tornado, pre-Jon Stewart…*). The suffix *-ness* which can be applied to adjectives to create nouns can also be applied productively (*crumbliness dingbattiness),* even though the productive suffix has a number of exceptions (*??youngness; ??honestness*). Yang suggests that productivity is determined by the following three key numbers:

1. the number of cases which potentially follow a rule: N
2. the number of witnessed exceptions to a rule: e
3. the number of witnessed rule-following cases: M

In particular, Tolerance and Sufficiency principles are claimed to provide a ceiling on the number of exceptions, and a floor on the number of cases witnessed following a rule (14):

Tolerance Principle (TP): $e \leqslant N/\ln N$
Sufficiency Principle (SP): $M \geq N - N/\ln N$ (equivalently, $N - M \leq N/\ln N$)

Challenges to the Tolerance Principle are addressed elsewhere (see Kapatsinski, this volume; Goldberg, in press), so this brief note will focus on the Sufficiency Principle (SP), which I argue is unrealistically demanding. The SP specifies the number of cases that must be witnessed following a rule (M) before the rule can be extended for use with other words. Yang (2016, p. 177) clarifies, "Before the positive evidence is sufficient – when M sits below the sufficiency threshold – learners lexicalize all M items and does [sic] not generalize beyond them."

https://doi.org/10.1075/lab.18061.gol
Linguistic Approaches to Bilingualism 8:6 (2018), 727–732.
ISSN 1879-9264 | E-ISSN 1879-9272

To be concrete, if a rule potentially applies to, say, 100 cases, the TP allows up to 22 of them to be exceptions; the SP then requires that fully 78 of them must be witnessed following the rule before it can be used productively (78 = 100 – 22). "Only when M crosses the Sufficiency threshold does R become a truly productive rule" (2016, p. 178). *PoLP* further assumes that learners may mistakenly make a rule productive, resulting in overgeneralization errors, but that they "can still backtrack to lexicalization if the amount of positive evidence [M] drops below the Sufficiency threshold" (2016, p. 213). That is, if a child recognizes that the rule may potentially apply to more cases than previously thought, say 500 cases ($N = 500$), *PoLP* predicts that the child will tolerate up to 80 exceptions but must witness at least *420* rule-following cases. Note that in order to count to 420, all rule-following instances must be retained throughout language learning. Each new rule-following case must be compared against each previously encountered rule-following case in order to know if the total number of rule-following cases should be increased.

Whenever the number of exceptions to a rule reaches its maximum, as it often does in the examples cited, *learners must witness and retain all other cases* that *potentially* follow a rule actually following the rule in order for the rule to become 'productive.' While children have been argued to be conservative learners, the SP takes conservatism to a whole new level.

If we assume that at the point when speakers know they can apply *pre-* productively, they have witnessed 1000 proper names or nouns that can potentially be construed to evoke a temporal onset, the SP requires that they must have actually witnessed roughly *855* distinct names or nouns being used with the *pre-* prefix. Note that SP requires that this enormous number must be witnessed *even though there are 0 exceptions*, because the SP does not make any reference to the number of witnessed exceptions (e). Surely this sets the cost of productivity unnecessarily high. Moreover, the need to retain massively long lists of rule-following cases before a rule becomes productive undermines the stated reason that a productive rule is created in the first place, as productive rules are assumed to increase efficiency: "learners postulate a productive rule only if it results in a more efficient organization of language, as measured in processing time, *rather than listing everything in lexical storage*" (2016, p. 9). But the Sufficiency Principle presumes that children retain all rule-following cases (as well as all exceptions) on an ongoing basis.

Fortunately, there are other ways to address the partial productivity puzzle (Ambridge et al., 2018; Barddol, 2008; Booij, 2018; Goldberg, in press; Kapatsinki, 2018; Zeschel 2012). It is important to bear in mind that productive formulations are created in order to satisfy semantic and discourse demands: that is, on the basis of communicative need. For instance, we might creatively coin a word like *pre-Trump* in order to identify a general time period in political life that has no other conventional name. We confidently apply the term because as we have witnessed other proper names being used in parallel ways (e.g., *pre-Watergate, pre-Columbine*). Goldberg (in press) argues that productivity is determined by the same inductive generalizations that are required to form the 'rules' (constructions) in the first place. The proposal also takes into account the fact that constructions compete with one another to express our intended messages. That is, when there exists a conventional alternative way to express our intended meaning, constructions that are otherwise productive are constrained. For example, the *-ness* ending cannot be applied when a conventional means of expressing the intended meaning already exists. For example, the words ?*youngness*, and ?*jealousness* are preempted or blocked by *youth* and *honesty* (Aronoff, 1976; Kiparsky, 1982).

The proposal in Goldberg (in press) is intended to apply to grammar, morphology, and word meaning. Previously witnessed partially-abstracted exemplars cluster together in our hyper-dimensional representational space for language, forming a massively interrelated dynamic system (a construct-i-con), which is an expanded version of the lexicon. We use whichever combination of constuctions is sufficiently accessible and best matches our intended message-in-context. The following points constitute the heart of the proposal:

- Speakers balance the need to be expressive and efficient while obeying the conventions of our speech communities.
- Our memory is vast but not perfect: memory traces are retained but partially abstract ('lossy').
- Lossy memories are aligned when they share relevant aspects of form and function, resulting in overlapping, emergent clusters of representations: Constructions.
- New information is related to old information (memory is associative), resulting in a rich network of constructions.
- During production, multiple constructions compete with one another to express our intended message.
- During comprehension, mismatches between what is expected and what is witnessed fine-tune our network of learned constructions via error-driven learning.

Goldberg (in press) also addresses age effects and differences between L1 and L2 learning. Research suggests two key factors lead to difficulty in reaching native-like proficiency in an L2, beyond the amount and type of input that L2 learners receive. The first is a subtle warping of the conceptual space that is used for "thinking for speaking" (Slobin, 1996). As adults, we have become highly practiced in the linguistic skills we already use regularly, and these skills constitute ingrained linguistic habits in which we use certain forms to express certain types of messages in certain types of contexts (chapter 4). We have implicitly learned that certain dimensions and not others are important for clustering exemplars used to express various types of messages-in-context. This is the case for sounds, words, lexically filled constructions, and abstract argument structure constructions. Spanish speakers use a plural definite determiner in 'generic' contexts (*Los perros son mamíferos*), native-English speakers use bare plurals (*Dogs are mammals*) in this context. Notably, Spanish speakers who learn English as a second language are prone to using the *English* definite determiner in generic contexts (?*The dogs are fun)* (Ionin & Montrul, 2010). L2 errors provide ample evidence for the idea that learning a language involves learning which constructions to use *in which contexts*.

Adult native speakers have assigned each particular construction of their L1 to a particular range of context types, and this assignment has been reinforced and fine-tuned over decades. Once a collection of context-types has been categorized together for the sake of an L1 construction, it becomes more difficult to assign an overlapping but distinct range of context types to a construction in L2.

A second difference is a reduction in competition-driven learning in L2, stemming from the added cognitive demands of using a second language. Native speakers are adept at generating expectations about upcoming words and forms as they comprehend language (e.g., Arnold et al., 2000; Dahan et al., 2000), and our expectations become fine-tuned through the process of error-driven learning (*statistical preemption*, Goldberg, chapter 5). That is, if we expect one thing and witness another, the error signal leads to a change in the strengths of the connections that predict which constructions are used in which contexts. This in turn leads to more accurate predictions in the future. A number of related findings suggest that L2 speakers are less likely than native speakers are to predict upcoming forms during online comprehension, even when they demonstrate knowledge of the forms during production and in off-line tasks (e.g., Grüter et al., 2014; Kaan et al., 2014; Lew-Williams & Fernald, 2010). To the extent that L2 learners' ability to predict upcoming grammatical forms is reduced, competition-driven learning will be correspondingly reduced. In particular, if non-native speakers do not anticipate upcoming utterances to the same extent as native speakers do, they will have less opportunity to learn from predictions that are subsequently corrected.

To summarize, attention to communicative needs, context, prior learning, and cognitive load are all necessary to account for when learners use constructions productively and how L2 speakers differ from native speakers. It is not sufficient to simply count numbers, and it is not necessary to count all potential instances.

References

Ambridge, B., Barak, L., Wonnacott, E., & Bannard, C. (2018). Effects of Both Preemption and Entrenchment in the Retreat From Verb Overgeneralization Errors: Four Reanalyses and an Extended Replication. *Collabra: Psychology* 4:1. https://doi.org/10.1525/collabra.133

Arnold, J. E., Eisenband, Brown-Schmidt S., & Trueswell, J. C. (2000). The Rapid Use of Gender Information: Evidence of the Time Course of Pronoun Resolution from Eyetracking. *Cognition* 76(1): 13–26. https://doi.org/10.1016/S0010-0277(00)00073-1

Aronoff, M. (1976). *Word formation in generative grammar*. Cambridge, Mass: Linguistic Inquiry Monographs.

Barðdal, J. (2008). *Productivity: Evidence from case and argument structure in Icelandic*. John Benjamins Publishing. https://doi.org/10.1075/cal.8

Booij, G. (2018). *The Construction of Words: Advances in Construction Morphology*. Springer.

Dahan, D., Swingley, D., Tanenhaus, M. K., & Magnuson, J. S. (2000). Linguistic Gender and Spoken-Word Recognition in French. *Journal of Memory and Language* 42(4): 465–80. https://doi.org/10.1006/jmla.1999.2688

Goldberg, A. E. (in press). *Explain me this: creativity, competition and the partial productivity of constructions*. Princeton: Princeton University Press.

Grüter, T., Hurtado, N., Marchman, V. A., & Fernald, A. (2014). Language Exposure and Online Processing Efficiency in Bilingual Development. *Input and Experience in Bilingual Development* 13: 15–36.

Ionin, T. & Montrul, S. (2010). The role of L1 transfer in the interpretation of articles with definite plurals in L2 English. *Language Learning*, 60(4), 877–925. https://doi.org/10.1111/j.1467-9922.2010.00577.x

Kaan, E. (2014). Predictive Sentence Processing in L2 and L1: What Is Different? *Linguistic Approaches to Bilingualism* 4(2): 257–82. https://doi.org/10.1075/lab.4.2.05kaa

Kapatsinski, V. (this volume). On the intolerance of the Tolerance Principle. *Linguistic Approaches to Bilingualism*, 8(6), 738–742.

Kapatsinski, V. (2018). *Changing minds changing tools: From learning theory to language acquisition to language change*. Cambridge: MIT Press.

Kiparsky, P. (1982). Word-formation and the lexicon. Mid-America Linguistics Conference.

Lew-Williams, C., & Fernald, A. (2010). Real-Time Processing of Gender-Marked Articles by Native and Non-native Spanish Speakers. *Journal of Memory and Language* 63(4): 447–64. https://doi.org/10.1016/j.jml.2010.07.003

Slobin, D. (1996). From 'Thought and Language' to 'Thinking for Speaking'. In J. J. Gumperz and S. C. Levinson (Eds.), *Rethinking Linguistic Relativity*, 70–96. Cambridge, UK: Cambridge University Press.

Yang, C. (2016). *The price of linguistic productivity: How children learn to break the rules of language*. Cambridge, MA: The MIT Press.

Yang, C. (2018) A formalist approach to language acquisition. *Linguistic Approaches to Bilingualism*, 8(6), 665–706.

Zeschel, A. (2012). *Incipient productivity: A construction-based approach to linguistic creativity* (Vol. 49). Walter de Gruyter.

Address for correspondence

Adele E. Goldberg
Department of Psychology
Princeton University
327 Peretsman Scully Hall
Princeton, NJ 08540
USA

adele@princeton.edu
https://orcid.org/0000-0001-8636-8890

Publication history

Date received: 3 July 2018
Date accepted: 23 August 2018

Mechanistic formal approaches to language acquisition

Yes, but at the right level(s) of resolution

Stefan Th. Gries
University of California, Santa Barbara

I am coming to Yang's target article (Yang, 2018) as a usage-based quantitative corpus linguist with interests, though not the greatest degree of expertise, in first/second/foreign language acquisition. There is a lot to like in the keynote paper both in terms of theoretical and methodological orientation: I am quite sympathetic to how Yang defines his formalist ('mechanistic') perspective; I appreciate his plea for methodological rigor and its emphasis over mode of explanation; I agree with his view that a (formal) theory of language acquisition ought to be more than a description of patterns and correlations but needs to include a mechanistic and causal account. Somewhat more specifically, I also particularly welcome the notion to develop mathematical models of language acquisition/learning especially when such models are probabilistic, can accommodate individual variation, and are compatible with a lesser role of innate parameters than much more generative work has been arguing for.

All that being said, it is a truism that the devil is in the details. Space does not permit me to discuss a variety of (admittedly smaller) issues so I will focus here on Yang's characterization of the state and standards of research on usage-based learning. In particular, Yang argues that "the evidence for an initial item-based stage of child language has been overstated" (p. 674), that claims regarding 'statistical dominance' are insufficient and claims regarding item-based learning 'require more work', that at a minimum, one needs to show that, say, unanalyzed collocations such as *give me* are conspicuously more frequent than a productive rule account would predict (p. 674).

But there are two reasons why his points of critique are not all compelling. First, Yang's terminology is just as imprecise as that of the usage-based work he discusses critically: what does 'statistically predominant' mean, what is 'conspicuously more

https://doi.org/10.1075/lab.18050.gri
Linguistic Approaches to Bilingualism 8:6 (2018), 733–737.
ISSN 1879-9264 | E-ISSN 1879-9272

frequent'? Even, or maybe especially, in a target article, we need more precise criteria – if only to agree with them or to admit that usage-based linguistics has not reached those thresholds. Relatedly, the choice of works cited is selective: Unless I missed something, with the exception of Pine et al. (2013), none of the usage-based original research articles in Section 3 'Rules vs. Storage', for instance, is less than 15 years old and even Pine et al. is only mentioned in a footnote but not where the issue of determiner imbalance is discussed in the main text. While I myself have issues with some of the usage-based acquisition work, this does not seem to compare both approaches on the same level of evolution; I would be very interested to see how Yang's variational model and equations stack up against, or even just compare and relate to, the kinds of studies discussed in Chapter 5 of Christiansen and Chater (2016) that highlight the power of distributional information or, for instance, work using the Vapnik-Chervonenkis dimension, which Christiansen and Chater (2016, 155f.) state "establishes an upper bound for the number of examples needed by a learning process that starts with a set of hypotheses about the task solution" and which Yang is of course familiar with.

The second reason why I feel that some of Yang's discussion falls a bit short is more important, however, and it is actually a reason for me to also be somewhat unhappy with some of the usage-based corpus-linguistic work on first/second language acquisition. Simply put, much like several of Yang's arguments and mathematical models, a lot of corpus-linguistic work severely underutilizes the huge amount of probabilistic information corpora have to offer by restricting itself to absolute and relative frequencies of (co-)occurrence of types and tokens (and maybe their ratios). However, as especially Nick Ellis and collaborators have argued repeatedly (e.g. Gries & Ellis, 2015; Ellis, 2016), much more information needs to be included: *frequency*, yes, but also *dispersion* of elements in a corpus (because dispersion is related to learning, Ambridge et al., 2006; Gries, 2008), *association/contingency* (Ellis, 2006), *predictability/surprisal* (Smith & Levy, 2013), *entropy, salience, prototypicality* (Ellis, Römer, & O'Donnell, 2016), and more. Thus, while I am genuinely sympathetic to the general idea of the variational model and the Tolerance/Sufficiency principles, these, too, try to explain something as complex as learning a first/second language – with all the multidimensional information that entails – on the basis of little more than type and token frequencies, which is certain to underestimate both the complexity of the task at hand and disregards any information regarding the salience or surprisal of exceptions to the hypothesized rules/grammars.

As a brief example, consider *gimme*, which Yang discusses in Section 3.1. First, it is not only just the high frequency of *give me* that led usage-based linguists to hypothesize that this might be a single unit but precisely the fact that it is frequently pronounced in the reduced version of *gimme*. Second, whether or not *give me*/*gimme* is a unit requires more 'methodological rigor' and clarity than a mere comparison of ratios of *give me*: *give him*: *give her* vs. *me*: him: *her*; there is much corpus-linguistic work on assessing association/contingency and I briefly checked the associations of the words following *give* in the Brown (1973) corpus (downloaded from <https://childes.talkbank.org/data/Eng-NA/Brown.zip> 10 June 2018).

These collocates were identified (rather heuristically) by loading and conflating each file into one string, breaking it up at utterance tiers, retrieving utterances containing v|give in the %mor tier, and finding sequences of non-spaces after *give* separately for children and caretakers. Then, I determined for each type attested after *give* its overall frequency in the corpus (separately for children and everyone else). Then, to keep effects of frequency and association separate (see Gries & Ellis, 2015; Gries, 2018), I collected for each collocate its frequency after *give* and the log odds ratio, an association measure ranging from $-\infty$ to $+\infty$, where the sign of the log odds ratio indicates whether the collocate is more often (positive) or less often (negative) observed after *give* than expected by chance; the absolute value of the log odds ratio quantifies the strength of the effect. To illustrate this approach for the word *her*, the data amount to those shown in Table 1, from which one can compute an odds ratio of $({}^{12}/_{300}\ /\ {}^{380}/_{89144}) \approx 9.3836$, leading to a log odds ratio of ≈ 2.239.

Table 1. 2×2 co-occurrence matrix for *give* and *her* in the Brown (1973) corpus

	Give	**Other words**	**Sum**
her	12	380	392
other words	300	89144	89444
Sum	312	89524	89836

The two panels of Figure 1 show the results for all collocates following *give*. As indicated in the *x*-axis labels, the left panel shows the results for all three children combined whereas the right one shows the results for the caretakers. In each panel, the *x*-axis represents the co-occurrence frequency of the collocate with *give* (logged to the base of 2) and the *y*-axis represents the log odds ratio of the collocate and *give* as computed above; lemmas printed in green exhibited a significant attraction/repulsion to the slot after *give*, lemmas in purple did not.

Clearly, the data are at least compatible with the notion that *gimme* might be a unit. Not only do we find that it is often transcribed with the phonological reduction from *give me* to *gimme*, but we also see in both the children's and the caretakers' output that *me* is the most frequent collocate after *give* and, more importantly, it is also the first or second most strongly attracted collocate to the

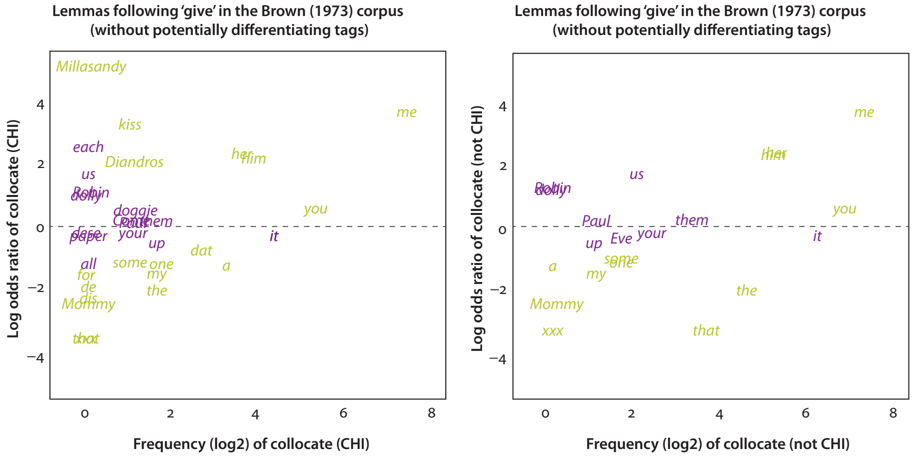

Figure 1. Frequencies and association measures for lemmas following *give* in the Brown (1973) corpus of language acquisition (left panel: children, right panel: caretakers)

position after *give* when the collocates overall frequencies in the corpus are taken into consideration. Similarly, *give me* is also the most evenly dispersed *give+x* collocation in the corpus (as measured by the *DP* index, see Gries, 2008), which means children's chance to encounter it (often) are highest; this is particularly relevant given how little of the actual input to and output of the child we actually have (Tomasello & Stahl, 2004).

None of this reduces the onus on usage-based work in general to provide "rigorous" testing of their L1/L2 acquisition data – I have been making similar points for years – but it shows that the situation may not be as clear cut as Yang makes it appear to be. While the above analysis is not certainly not the most sophisticated corpus-linguistic analysis one can do (and of course it's also no proper linguistic analysis), it shows that a mere comparison of frequency ratios is neither. A position paper that criticizes usage-based work for 'overstating their evidence,' not doing 'enough work,' and lacking proper hypotheses/rigor needs to be more compelling even if much usage-based work has also not done the ideal corpus statistics: All sides of the debate – formalists, usage-based linguists, but also general corpus linguists like myself – need to step up their quantitative corpus-linguistic game.

References

Ambridge, B., Theakston, A. L., Lieven, E. V. M., & Tomasello, M. (2006). The distributed learning effect for children's acquisition of an abstract syntactic construction. *Cognitive Development*, 21(2), 174–193. https://doi.org/10.1016/j.cogdev.2005.09.003

Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press. https://doi.org/10.4159/harvard.9780674732469

Christiansen, M. H. & Chater, N. (2016). *Creating language: integrating evolution, acquisition, and processing*. Cambridge, MA: The MIT Press.

Ellis, N. C. (2006). Language acquisition as rational contingency learning. *Applied Linguistics*, 27(1), 1–24. https://doi.org/10.1093/applin/ami038

Ellis, N. C. (2016). Cognition, corpora, and computing: triangulating research in usage-based language learning. *Language Learning*, 67(51), 40–65. https://doi.org/10.1111/lang.12215

Ellis, N. C., Römer, U., & O'Donnell, M. B. (2016). *Usage-based approaches to language acquisition and processing*. New York: Wiley-Blackwell.

Gries, St. Th. (2008). Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics*, 13(4), 403–437. https://doi.org/10.1075/ijcl.13.4.02gri

Gries, St. Th. (2018). Operationalizations of domain-general mechanisms cognitive linguists often rely on: a perspective from quantitative corpus linguistics. In S. Engelberg, H. Lobin, K. Steyer, & S. Wolfer (Eds.), *Wortschätze: Dynamik, Muster, Komplexität* (pp. 75–90). Berlin & Boston: De Gruyter. https://doi.org/10.1515/9783110579963-005

Gries, St. Th. & Ellis, N. C. (2015). Statistical measures for usage-based linguistics. *Language Learning*, 65(Supplement 1), 1–28. https://doi.org/10.1111/lang.12119

Pine, J. M., Freudental, D., Krajewskyi, G., & Gobet, F. (2013). Do young children have adult-like syntactic categories? Zipf's law and the case of the determiner. *Cognition*, 127(3), 345–360. https://doi.org/10.1016/j.cognition.2013.02.006

Smith, N. J. & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, 128(3), 302–319. https://doi.org/10.1016/j.cognition.2013.02.013

Tomasello, M. & Stahl, D. (2004). Sampling children's spontaneous speech: how much is enough? *Journal of Child Language*, 31(1), 101–121. https://doi.org/10.1017/S0305000903005944

Yang, C. (2018). A formalist perspective on language acquisition. *Linguistic Approaches to Bilingualism*, 8(6), 665–706.

Address for correspondence

Stefan Th. Gries
Department of Linguistics
University of California, Santa Barbara
Santa Barbara, CA 93106–3100
United States of America

stgries@linguistics.ucsb.edu
https://orcid.org/0000-0002-6497-3958

Publication history

Date received: 11 June 2018
Date accepted: 15 August 2018

On the intolerance of the Tolerance Principle

Vsevolod Kapatsinski
University of Oregon

Yang (2018) argues for renewing our commitment to mechanistic theories of language and of language acquisition, which he (somewhat surprisingly) calls "formalism". As a usage-based linguist, I have no objection to mechanistic explanation, which has been a hallmark of both usage-based linguistics and of connectionist approaches to psychology and neuroscience (e.g., Bybee & McClelland, 2005). Similarly, deductive reasoning, where an "intuitive grasp of the essentials or a large complex of facts leads the scientist to the postulation of a hypothetical basic law" (Einstein, 1919) has long been at the core of cognitive science, which derives predictions either from how the brain works (connectionism) or from functional analyses of the task facing the learner (rational analysis).

However, deductive reasoning is only as good as the axioms it assumes and the consistency of its application. A great strength of Yang's Variational Model is that it is grounded in classic learning theory. In contrast, the Tolerance Principle (TP) lacks any plausible grounding. The basic assumption behind TP is that a rule is applied after the learner searches through a list of exceptions to the rule, serially, in inverse order of frequency. The serial search assumption is crucial for the claim that TP minimizes processing time, and is adopted from Forster (1976). Despite Forster's otherwise enormous influence on psycholinguistics, serial search has never been accepted by the field. Instead, every single other model of lexical processing has adopted an alternative approach, cued retrieval through parallel activation, e.g., Morton (1969). The unpopularity of serial search is due to the fact that it is incompatible with the parallel nature of the brain. It is a basic neuroscientific fact that a representation of something as complex as a spoken word requires a large population of neurons. Each neuron also participates in representing other words: neural representations are 'distributed'. A serial search through such a system is physically impossible because activating any one of the neurons in question partially activates multiple words. Without serial search, there is no motivation for the TP.

Extraordinary claims require extraordinary evidence. What then is the "large complex of facts" that supports serial search? There is only one, which is that rank

https://doi.org/10.1075/lab.18052.kap
Linguistic Approaches to Bilingualism 8:6 (2018), 738–742.
ISSN 1879-9264 | E-ISSN 1879-9272

frequency is a slightly better predictor of reaction times and errors in lexical decision tasks compared to log frequency. Yang's data for irregular verbs show $R^2 = 67\%$ vs. 64% (2016, p. 51). However, the superiority of the rank transformation is likely to be a statistical artifact (Adelman & Brown, 2008; Rouder et al., 2008). Rank transformations are maximally robust to outliers and differences in scale and distribution shape between the correlated variables. The fact that Spearman regression fits better than Pearson regression says little about the process that gave rise to the dataset. Other findings claimed to be problematic for parallel, associative models of lexical processing by Yang (2016, p. 52) have actually provided some of the most convincing evidence for such models in prior work (Ramscar et al., 2014). We await any behavioral evidence in lexical processing that supports serial search over alternatives that are consistent with neuroscience.

It is also impossible to justify TP based on rational analysis of the inference task facing the learner. Consider the example of deciding whether a new species will have a characteristic that applies to 8 of the known species inhabiting a particular island or the one that applies to 2 of the species (Yang, 2018, p. 681). The decision will obviously depend not only on the number of species in each cluster but on how far the new species is from the two vs. the eight, how similar the existing species in each cluster are to each other, and how certain one is that each species does in fact possess the characteristics in question. Trying to reduce the decision to the number of species that do and do not have the characteristic is a hopeless endeavor that no statistician would seriously consider attempting. Are language learners really attempting to do the impossible?

In addition to being based on sound, well-supported assumptions, a good deductive hypothesis should be applied consistently. One inconsistency in Yang's paper concerns the use of indirect negative evidence. Yang argues that absence of evidence is not evidence of absence and therefore the use of indirect negative evidence is a fallacy. However, Yang's Sufficiency Principle (SP) proposes that items that are not known to undergo a rule behave exactly like exceptions to the rule (see also Goldberg, this volume). Under SP, verbs whose past tense forms have not been observed are assumed to provide evidence against the productivity of the 'add *-ed*' rule. Because SP treats absence of evidence in exactly the same way TP treats evidence of absence, the two principles together are an example of the kind of logic Yang argues to be a fallacy. Note that a child acquiring morphology cannot use only TP or only SP because s/he both knows some exceptions and has no information about other items. Yet, Yang proceeds to use only SP or only TP depending on example. If one used both, the sum of items not known to obey the rule would be much greater, potentially rendering even regular rules in rare paradigm cells unproductive.

Yang (2018, 676–678) notes that both TP and SP – if applied to all nouns in the input witnessed by a child – fail to predict that a noun can usually take both determiners. The less frequent nouns often occur with only one determiner in the input. He concludes that the child learns that determiners are interchangeable on the basis of a few frequent nouns and then sticks with this decision as more nouns are learned. However, the same child is assumed *not* to stick with the decision that the ditransitive and the prepositional dative are interchangeable (Yang, 2018, 684–686). The child learns that possession verbs occur with both constructions, but then abandons this generalization with more input, retreating from over-generalization. How does the child know whether to retreat when his current SP/TP values tell him to retreat but his early experience tells him to persist?

There are also empirical issues with the predictions of TP. Yang proposes that the learner rejects grammars that do not meet TP and then partitions the lexicon into subsets that support TP-following grammars. If more than one grammar exceeds TP, then the more reliable grammar is chosen (2016, p. 97). At least two empirical problems arise (see Kapatsinski, 2018, for more). First, there is no way for an exceptionless rule to lose productivity. Second, there is no way to account for extension of multiple competing grammatical patterns to new words in a probabilistic manner. Both of these are, of course, common diachronic phenomena.

For example, loanword adaptation data show that Russian velar palatalization is fully productive before the suffixes *-ek* and *-ok* but only partially productive before the stem extension *-i-* (Kapatsinski, 2010). Yet, there are zero exceptions to velar palatalization in any of these contexts in the established lexicon. Given a traditional generative description of Russian velar palatalization, these data are incompatible with TP.

The low productivity of palatalization before *-i-* does follow if rules attaching stem extensions other than *-i-* can compete with rules that attach *-i-* and palatalize. In that case, the palatalizing rules are no longer free of exceptions. Productivities of such rules in loanword data turn out to closely track their reliabilities in the native lexicon. However, if rule induction were guided by TP, these rules should never have been extracted, as they do not minimize rule competition.

Furthermore, according to Yang, rules are either completely productive or completely unproductive. However, productivity of palatalization before *-i-* varies from ~90% to ~70% to ~30% to ~10% across segmental contexts, tracking rule reliability differences across contexts in a gradient manner. Individuals exposed to miniature artificial languages modeled on Russian also show clear evidence of gradient productivity with individual palatalization probabilities spanning the full range from 0% to 100% (Kapatsinski, 2010). Where does one draw the line between productivity and lack thereof?

In summary, the Tolerance Principle does not solve the problems it is claimed to solve. It provides no mechanism for inducing the generalizations human learners infer from the primary linguistic data they experience, and it is not applied to data consistently. However, the root problem is that the starting assumptions TP is derived from are supported by no independent evidence and contradict the most basic principles of cognitive neuroscience. While I applaud Yang's call for a mechanistic approach to language acquisition grounded in independent principles, the Tolerance Principle does not fit the bill.

References

Adelman, J. S., & G. D. A. Brown. (2008). Modeling lexical decision: The form of frequency and diversity effects. *Psychological Review*, 115(1), 214–227. https://doi.org/10.1037/0033-295X.115.1.214

Bybee, J., & J. L. McClelland. (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review*, 22(2–4), 381–410. https://doi.org/10.1515/tlir.2005.22.2-4.381

Einstein, A. (1919). Induction and deduction in physics. *Berliner Tageblatt*, (December 25).

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

Kapatsinski, V. (2018). Words versus rules (Storage versus on-line production/processing) in morphology. In M. Aronoff (Ed.), *Oxford Research Encyclopedia of Linguistics*. Oxford: Oxford University Press.

Kapatsinski, V. (2010). Velar palatalization in Russian and artificial grammar: Constraints on models of morphophonology. *Laboratory Phonology*, 1(2), 361–393. https://doi.org/10.1515/labphon.2010.019

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, 76(2), 165–178. https://doi.org/10.1037/h0027366

Ramscar, M., P. Hendrix, C. Shaoul, P. Milin, & H. Baayen. (2014). The myth of cognitive decline: Non-linear dynamics of lifelong learning. *Topics in Cognitive Science*, 6(1), 5–42. https://doi.org/10.1111/tops.12078

Rouder, J. N., F. Tuerlinckx, P. Speckman, J. Lu, & P. Gomez. (2008). A hierarchical approach for fitting curves to response time measurements. *Psychonomic Bulletin & Review*, 15(6), 1201–1208. https://doi.org/10.3758/PBR.15.6.1201

Yang, C. (2016). *The price of linguistic productivity: How children learn to break the rules of language*. Cambridge, MA: MIT Press.

Yang, C. (2018). A formalist perspective on language acquisition. *Linguistic Approaches to Bilingualism*, 8(6), 665–706.

Address for correspondence

Vsevolod Kapatsinski
Department of Linguistics
University of Oregon
1290 University of Oregon
Eugene, OR 97403
United States of America

vkapatsi@uoregon.edu

Publication history

Date received: 13 June 2018
Date accepted: 15 August 2018

The importance of input representations

Jeffrey Lidz and Laurel Perkins
University of Maryland

Language learners use the data in their environment in order to infer the grammatical system that produced that data. Yang (2018) makes the important point that this process requires integrating learners' experiences with their current linguistic knowledge. A complete theory of language acquisition must explain how learners leverage their developing knowledge in order to draw further inferences on the basis of new data. As Yang and others have argued, the fact that input plays a role in learning is orthogonal to the question of whether language acquisition is primarily knowledge-driven or data-driven (J. A. Fodor, 1966; Lidz & Gagliardi, 2015; Lightfoot, 1991; Wexler & Culicover, 1980). Learning from data is not incompatible with approaches that attribute rich initial linguistic knowledge to the learner. On the contrary, such approaches must still account for how knowledge guides learners in using their data to infer a grammar.

Yang discusses three computational models that together paint a fuller picture of how learners use their input in grammar acquisition. But this picture is still missing a critical piece: how the learner's input is perceived, and what role immature perceptions play in the learning process. We argue that Yang's strategy of abstracting away from the learner's representation of the input leaves out the role of development, and as a result, may miss important generalizations about language acquisition.

Yang's variational model addresses one way in which language acquisition is incremental: children must learn from one sentence at a time. This model describes how learners might update their probabilities over possible grammars as they receive their input sequentially. But the incremental nature of language acquisition runs deeper. The way that learners perceive their input changes as their grammatical knowledge develops, and these perceptions determine what can and cannot be learned (Gagliardi & Lidz, 2014). For example, before children can identify the category "verb" in their input, that input is not useful for evaluating whether their language has verb raising. But learning cannot wait until children can veridically parse every sentence they hear; in this case no learning would be

https://doi.org/10.1075/lab.18051.lid
Linguistic Approaches to Bilingualism 8:6 (2018), 743–748.
ISSN 1879-9264 | E-ISSN 1879-9272

necessary (J. D. Fodor, 1998; Valian, 1990). Instead, children must parse their input as best they can with their developing linguistic knowledge, and those partial and immature parses are the input to a learning mechanism. This raises the question of how learners avoid being misled by incomplete or incorrect representations of their input.

This issue can be seen in the following example. Infants make inferences about verbs' meanings and argument structure on the basis of observing how they distribute in transitive and intransitive clauses (Fisher, Gertner, Scott, & Yuan, 2010; Lidz, White, & Baier, 2017). But reliably perceiving those distributions is not trivial, given the variability in how transitivity can be realized:

(1) Amy fixed her bicycle.

(2) * Amy fixed.

(3) What did Amy fix?

Recognizing that (3) is underlyingly transitive depends on knowing that *what* acts as the verb's object, despite not being realized in an argument position. A child who does not yet know that *what* is a *wh*-word might fail to represent it as an argument and treat the verb as intransitive. This could lead to faulty inferences about the argument structure and meaning of *fix*: a learner might think that this verb can freely occur without a direct object.

In considering the problem that clause types like *wh*-questions pose for grammar learning, Pinker (1984) proposed that learners must somehow filter out these sentences at stages of development when they cannot parse them accurately. This introduces a separate problem of how learners know which sentences to filter out. Perkins, Feldman, & Lidz (2017) demonstrated that a learner that expects error in its sentence representations can learn how much data to filter in order to identify the transitivity of verbs in child-directed speech, without knowing in advance which sentences were parsed inaccurately.

Thus, in order to ask how a learner's input provides relevant evidence for evaluating grammatical properties, we must ask how that input is perceived at the relevant point in development. Yang proposes that certain types of evidence would allow a learner to disambiguate between different grammars of argument drop (example 18, p. 688). For example, null objects are an unambiguous cue for a Chinese-type grammar, and occur in high enough proportion in speech to Chinese-learning infants to enable the acquisition of topic-drop at an early age. But this analysis does not account for how infants come to represent null objects in sentences that contain them. It also does not explain how English-learning infants rule out null objects in sentences like (3). If English-learning infants consider an object-drop analysis for these sentences, they might take them as evidence for a topic-drop grammar. However, if learners use a mechanism like filtering to deal with expected error in their own sentence perceptions, both English- and Chinese-learning infants might be cautious about when they draw inferences from sentences that are missing objects and what inferences they draw.[1] In this case, sentences with missing objects – regardless of whether they actually contain object drop – might *not* be trusted as good evidence for inferring a grammar of argument-drop. Learners' developing perceptions of their input interact non-trivially with the way that they use that input for learning.

The case of dative generalization and retreats raises an extreme version of this concern about the child's perception of her input. Yang's analysis depends upon the accurate perception of the semantic properties of the verbs independent of their syntax, percepts that are not likely to be possible given the messy relation between event perception and linguistic description.

The problem under consideration is (a) what allows children to overgeneralize the use of the double object construction (DOC) and (b) what subsequently allows them to retreat from overgeneralization. According to the Tolerance Principle, learners will link the DOC to the semantics of caused possession. As they acquire more verbs indicating caused possession that do not occur in the DOC, the Sufficiency Principle will cause learners to retract the link between DOCs and caused possession.

This elegant solution depends on the learner having accurate perceptions of caused possession. But how can caused possession be tokened in events, so that learners can formulate the relevant hypothesis about how to subcategorize ditransitive verbs? Since the link between caused possession and the DOC is what is to be acquired, the learner cannot use the fact that a verb occurs in that construction as evidence for its meaning. Instead, the learner would have to use information about the events described by the sentences containing that verb. But, caused possession in the world looks a lot like change of location. If John throws the ball to Mary, it will usually be true that Mary comes into possession of the ball. However, *throw* only implies caused possession in the DOC. The conceptual perspective on the event is dependent on the sentence used to describe it. The same event in the world can be construed as a change of possession or a change of location. Similarly, whereas only *tell* involves caused possession, all telling events are also sayings, and

1. There might be an interesting connection between a learner's degree of trust in her input and the parameter γ in Yang's variational model. This parameter represents "the magnitude of probability adjustment as the result of analysis" (p. 669), which Yang attributes to individual extra-linguistic cognitive factors. But the degree to which learners update their grammatical beliefs on the basis of input may also depend on how much error they expect in their input representations, given their current linguistic knowledge.

the vast majority of sayings are also tellings. So, if the caused possession meaning component isn't uniquely identifiable by nonlinguistic perception, any learning theory based on that perception cannot be correct (Gleitman, 1990).

Second, this learning theory would allow for a greater variation than is found among the world's languages. While the theory identifies rule-like links between the DOC and caused possession, the very same system could acquire a language where the prepositional dative was associated with caused possession. But this is not how languages work. If a language has two ditransitive constructions, the one expressing caused possession is always the one in which the goal c-commands the theme (Harley, 2002). And, children know this link despite a severe poverty of evidence (Viau & Lidz, 2011).

The solution to the problem of identifying subcategories of ditransitive verbs fails in two respects. It provides the learner with unrealistic input, since the relevant meaning components are not identifiable based on observations of the world. And, it fails to take advantage of cross-linguistically stable properties that give the learner some purchase on matching sentences to interpretations.

Yang's work illustrates the critical importance of making precise models of how the input is taken in and used for updating grammatical representations. Such precise models, by requiring researchers to provide rigorous analyses of how input drives the growth of grammar, can help to narrow the divide between usage-based and nativist approaches. In building such models, it is important to provide realistic assessments of what children can represent in their linguistic and extralinguistic environments and how these representations feed forward for learning.

Funding

This work was supported by grants from the National Science Foundation BCS-1551629 awarded to Jeffrey Lidz, and National Science Foundation DGE-1449815 and BCS-1827709 awarded to Laurel Perkins.

References

Fisher, C., Y. Gertner, R. M. Scott, & S. Yuan. (2010). Syntactic bootstrapping. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(2), 143–149. https://doi.org/10.1002/wcs.17

Fodor, J. A. (1966). How to learn to talk: Some simple ways. In F. Smith & G. A. Miller (Eds.), *The genesis of language: A psycholinguistic approach* (pp. 105–122). Cambridge, MA: MIT Press.

Fodor, J. D. (1998). Parsing to learn. *Journal of Psycholinguistic Research*, 27(3), 339–374. https://doi.org/10.1023/A:1023255705029

Gagliardi, A., & J. Lidz. (2014). Statistical insensitivity in the acquisition of Tsez noun classes. *Language*, 90(1), 58–89. https://doi.org/10.1353/lan.2014.0013

Gleitman, L. R. (1990). The structural sources of verb meanings. *Language Acquisition*, 1(1), 3–55. https://doi.org/10.1207/s15327817la0101_2

Harley, H. (2002). Possession and the double object construction. *Linguistic Variation Yearbook*, 2(1), 31–70. https://doi.org/10.1075/livy.2.04har

Lidz, J., & A. Gagliardi. (2015). How nature meets nurture: Universal grammar and statistical learning. *Annu. Rev. Linguist*, 1(1), 333–353. https://doi.org/10.1146/annurev-linguist-030514-125236

Lidz, J., A. S. White, & R. Baier. (2017). The role of incremental parsing in syntactically conditioned word learning. *Cognitive Psychology*, 97, 62–78. https://doi.org/10.1016/j.cogpsych.2017.06.002

Lightfoot, D. (1991). *How to set parameters: Arguments from language change*. Cambridge, MA: MIT Press.

Perkins, L., N. H. Feldman, & J. Lidz. (2017). Learning an input filter for argument structure acquisition. In *Proceedings of the 7th Workshop on Cognitive Modeling and Computational Linguistics*. https://doi.org/10.18653/v1/W17-0702

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Valian, V. (1990). Logical and psychological constraints on the acquisition of syntax. In L. Frazier & J. G. De Villiers (Eds.), *Language processing and language acquisition*. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-011-3808-6_5

Viau, J., & J. Lidz. (2011). Selective learning in the acquisition of Kannada ditransitives. *Language*, 87(4), 679–714. https://doi.org/10.1353/lan.2011.0088

Wexler, K., & P. Culicover. (1980). *Formal principles of language acquisition*. Cambridge, MA.

Yang, C. (2018). A formalist perspective on language acquisition. *Linguistic Approaches to Bilingualism*, 8(6), 665–706.

Address for correspondence

Jeffrey Lidz
Department of Linguistics
University of Maryland
1401 Marie Mount Hall
College Park, MD 20742

jlidz@umd.edu
https://orcid.org/0000-0001-8829-1495

Co-author details

Laurel Perkins
Department of Linguistics
University of Maryland
1401 Marie Mount Hall
College Park, MD 20742

perkinsl@umd.edu

Publication history

Date received: 13 June 2018
Date accepted: 23 August 2018

Learning a second language takes more than math

Silvina A. Montrul
University of Illinois

The nature and development of a first language by young children, who receive little instruction and guidance from caregivers, is a tantalizing mystery. Both the human genetic endowment and linguistic experience work admirably well together to guide and shape the process and outcome of language acquisition to linguistic maturity. Yang's (2002) variational learning model happily marries the best of generative and usage-based approaches by modeling computationally how Universal Grammar has to rely on input. Yang argues that at the heart of this problem lie children's endowed mathematical abilities, which allow them to perform statistical analyses of the input data. For example, the Linear Reward Penalty scheme (Bush & Mosteller, 1951) presents competing options, in our case grammars or rules, and probabilities in the input either punish or reward one or the other. The grammar that is rewarded wins out and the loser is eventually abandoned. If the Linear Reward Penalty Scheme accounts for how grammars develop, the Tolerance Principle and its corollary – the Sufficiency Principle (Yang, 2016) – are also equations that explain how, in the absence of negative evidence, children discover productive rules and exceptions to a rule as a function of their vocabulary size. Granting that L2 learners are faced with the same logical problem as L1 learners and must also build grammatical representations from input, Yang (2018) suggests that if these same equations are operative in L2 acquisition they may inform about differences and similarities between child and adult language.

Yang and Montrul (2017) considered how the Tolerance and Sufficiency principles operate in the L2 acquisition of the English dative alternation (*John gave a book to Peter/John gave Peter a book*), and Yang explains in the keynote article how variational learning may apply to the L2 acquisition of null subject grammars. Yet, the question of why children are so successful at eventually mastering rules and exceptions of their language while L2 learners show variable degrees of success remains a core puzzle in formal and psycholinguistic approaches to L2 acquisition. As an example, we have evidence from several languages that the phenomenon

https://doi.org/10.1075/lab.18055.mon
Linguistic Approaches to Bilingualism 8:6 (2018), **749–752**.
ISSN 1879-9264 | E-ISSN 1879-9272

of Differential Object Marking (DOM) is easily acquired in L1 acquisition and children make very few errors (Avram, 2015). Yet, bilingual children, adult L2 learners, and heritage speakers show significant variability with this same phenomenon (Guijarro Fuentes, 2012; Montrul & Sánchez-Walker, 2013; Ticio, 2015). The acquisition of gender assignment and agreement in gender-system languages is another example (Grüter et al., 2012). Why are monolingual children so good, while bilinguals and adult L2 learners are less so?

Yang's solution is to invoke, and flesh out more clearly, Newport's (1990) Less is More Hypothesis: "by the virtue of having greater cognitive capacities, or perhaps due to the influence of their L1 lexicon in developing an L2 lexicon… adult learners may have too many words for their own good." (Yang, keynote article p. 693). The idea here is that young children work on smaller data sets and perhaps perform bottom up processing to discover the rules (they focus on the trees to get to the forest); adults, on the other hand, take on more information and may deploy primarily top down processes (they cannot see the trees for the forest), thus biting off more than they can chew. In principle, the notion of smaller versus larger lexicons is a viable possibility to account for L1 and L2 acquisition differences, but this idea by itself cannot explain all that needs to be explained in relation to L2 adults. There are several other factors involved in L2 acquisition but space limitations allow me to address only two. The first obvious difference is the presence of another language (the L1) and, depending on the age of the learners, the degree of entrenchment of this other language. Another major difference is the input itself: its modality (visual, auditory), its availability (how often, how frequent), its quality (native, non-native). The entrenched L1 and the input may shape input processing skills, which would be crucial to perform statistical computations of the input for rule generation and detection of exceptions. Not only is the representation of the L1 the initial hypothesis in L2 acquisition (Schwartz & Sprouse, 1996) in many instances, but it may also dictate how input becomes intake, and what aspects of the language L2 learners actually compute from the input. If the input cannot be perceived, or is perceived differently, because the linguistic representation of the L1 guides what is attended to and what is ignored, then the route of L2 acquisition and the endstate are likely to diverge. Not only is the input that L2 learners receive less abundant in quantity than the input an L1 learner receives (in terms of hours per day), but it also differs in quality. L2 learners, who are older than L1 learners, are exposed to different input than child L1 learners, especially if they acquire the language in an instructed setting and are introduced to reading and writing earlier than an L1-acquiring child. A barrier to further progress in research in L2 acquisition is that estimating, measuring, and quantifying L2 input is far more challenging than recording, transcribing and annotating child directed speech.

As Carroll (1999) reminded us, input is not just something "out there" but it is shaped by how it becomes intake (Krashen, 1981; VanPatten, 1996). The problem of language acquisition is not the poverty of the stimulus but the perception and the parsing of the stimulus feeding grammars. That L2 learners very often do not perceive and process phonological and morphological contrasts in the L2 that are not present in their L1 has been amply demonstrated (Finn & Hudson-Kam, 2008), so it is clear that the L1 can shape the aspects of the processing patterns of the L2. Suppose a child learner of a DOM language (Romanian, Spanish, Turkish, Hindi, etc.) learns the distribution of DOM from a data sample of 50 objects. An adult L2 learner of a DOM language whose L1 does not have DOM might find it more difficult to learn the distribution of DOM in the L2, not because their vocabulary sample is necessarily larger but because their L1 representations funnel processing skills to ignore animacy or specificity cues relevant for DOM in the linguistic input. A similar problem arises with the acquisition of gender. Here again, L1 children are pros, L2 learners are not. An influential proposal is that problems with gender agreement in L2 acquisition are related to gender assignment in the lexicon, and to the way young children and adult L2 learners process input (Grüter et al., 2012). Studies have also found that the learners whose L1 has gender are more successful at processing and acquiring gender in the L2 than those whose languages do not have gender, stressing again how previous linguistic knowledge guides input processing and, eventually, the kinds of computations and equations that L2 learners apply to build grammatical knowledge from input. To find continuity between L1 and L2 acquisition, future studies must rigorously control these and other factors in order to isolate and elucidate how the potential mathematical equations that might allegedly underlie both L1 and L2 acquisition actually operate in L2 acquisition.

References

Avram, L. (2015). The L1 acquisition of differential object marking. *Revue Roumaine de Linguistique (RRL)*, LX (4), 331–338. București.

Bush, R. & Mosteller, F. (1951). A mathematical model for simple learning. *Psychological Review*, 68(3), 313–323. https://doi.org/10.1037/h0054388

Carroll, S. (1999). *Input and evidence. The raw material of second language acquisition*. Amsterdam: John Benjamins.

Finn, A. & Hudson Kam, C. (2008). The curse of knowledge: First language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition*, 108, 477–499. https://doi.org/10.1016/j.cognition.2008.04.002

Guijarro Fuentes, P. (2012). The acquisition of interpretable features in L2 Spanish: Personal *a*. *Bilingualism, Language and Cognition*, 15(4), 701–720. https://doi.org/10.1017/S1366728912000144

Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2A: A production or a real-time processing problem? *Second Language Research*, 28(2), 191–216. https://doi.org/10.1177/0267658312437990

Krashen, S. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon.

Montrul, S. & Sánchez-Walker, N. (2013). Differential object marking in child and adult Spanish heritage speakers. *Language Acquisition*, 20(2), 109–132. https://doi.org/10.1080/10489223.2013.766741

Newport, E. (1990). Maturational constraints on language learning. *Cognitive Science*, 14(1), 11–28. https://doi.org/10.1207/s15516709cog1401_2

Schwartz, B. & Sprouse, R. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research*, 12(1), 40–72. https://doi.org/10.1177/026765839601200103

Ticio, E. (2015). Differential object marking in Spanish-English early bilinguals. *Linguistic Approaches to Bilingualism*, 5(1), 62–90. https://doi.org/10.1075/lab.5.1.03tic

VanPatten, B. (1996). *Input processing and grammar instruction*. Chestnut Hill, NJ: Ablex.

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

Yang, C. (2016). *The price of linguistic productivity: How children learn to break rules of language*. Cambridge, MA: MIT Press.

Yang, C. (2018). A formalist perspective on language acquisition. *Linguistic Approaches to Bilingualism*, 8(6), 665–706.

Yang, C. & Montrul, S. (2017). Learning datives: The tolerance principle in monolingual and bilingual acquisition. *Second Language Research*, 33(1), 119–144. https://doi.org/10.1177/0267658316673686

Address for correspondence

Silvina A. Montrul
University of Illinois
Dept. of Linguistics, MC-168
707 South Mathews Avenue
Urbana, IL 61801
United States of America

montrul@illinois.edu
https://orcid.org/0000-0001-6011-5959

Publication history

Date received: 20 June 2018
Date accepted: 15 August 2018

Language-level input factors are not enough to explain child bilingual acquisition

Johanne Paradis
University of Alberta

Yang's (2018) article is focused on illustrating how his Variational Model can account for some well-known phenomena in acquisition through a constrained formal system that is driven by input properties. For example, in Yang's model, crosslinguistic differences in the timing of children's acquisition of the parameter setting for obligatory subjects vs. topic-drop can be explained by language-specific differences in the distributional frequency of input properties needed to select the correct parameter. Yang writes that "…variational learning is gradual, probabilistic, and quantity sensitive" (Yang, 2018, p. 689) – a statement that underscores how his approach builds bridges between generative and usage-based approaches. Yang mainly focuses on input variation at the language level in illustrating the model, and does not systematically consider input variation at the level of the individual learner. And yet, individual differences in the quantity and quality of input experienced modulate access to – and uptake of – the language-level properties discussed in this keynote article. If this formalist model is to have useful applications to bilingual acquisition, the role of input variation at the child level must be as systematically incorporated as variation at the language level. This is because there are more complex sources of input variation at the child level in bilingual learners than in monolingual learners. It has been estimated that as many children grow up bilingual as monolingual across the globe (Grosjean, 2010); therefore, any model of child language acquisition that is input-driven should be able to account for the acquisition experience and outcomes of bilingual children.

How are language-level and child-level input variation defined? "Language level" refers to the input properties pertaining to the target language itself. For example, the frequency and distribution of morphemes, words, and collocations thereof, vary in the linguistic input to children. Yang's keynote article examines in painstaking detail how these language-level properties can be shown to drive rule formation, and in turn, how exceptions and irregularities in the input can be accounted for in a formal model of rule-learning and hypothesis competition.

https://doi.org/10.1075/lab.18059.par
Linguistic Approaches to Bilingualism 8:6 (2018), 753–757.
ISSN 1879-9264 | E-ISSN 1879-9272

Language-level input properties affect all learners of a target language. However, individual children vary in their access to, and uptake of, these language-level properties. Variation in individual bilingual children's linguistic environment is greater than that of monolingual children, and thus, there is even more modulation of access to the language-level properties for this population. Bilingual children have their input space divided between two languages; this division is seldom 50–50 and the relative quantity of input in each language can shift over time and according to social context (home vs. school). Bilingual children might be exposed regularly to L2 input from speakers with varying degrees of proficiency within their families. In the case of children from migrant families, the linguistic diversity and complexity of the societal L2 input could be very different from that of their heritage L1. Finally, bilingual children's onset of acquisition of their two languages can be staggered, which means they are older when access to L2 input begins than children acquiring the same language as an L1.

The relationship between individual variation in quantity of input and rate of acquisition has been well-established in the literature for both monolingual and bilingual children, e.g., a child receiving more input than another child typically shows faster acquisition. For bilinguals, relative quantity of input in each language interacts with language-level input properties, as found in a study we conducted on past tense acquisition in French-English bilinguals (Paradis, Nicoladis, Crago, & Genesee, 2011). Irregular past tense verbs in English are highly idiosyncratic and most schemas have zero to low type frequency; whereas, in French, irregular verbs usually cluster in families of schemas with moderate type frequency. Because rate of morphological rule acquisition is driven mainly by type frequency (Bybee, 2001; 2008), this would lead to the prediction that children would acquire irregular past tense in French faster than in English, which is what we found. For bilingual children, this relationship between language-level input properties interacted with individual variation in the relative amount of input a child received in each language. So, bilingual children who had predominantly French input at home acquired French irregulars in advance of their bilingual peers who had predominantly English input at home, and bilinguals with predominantly English input at home showed similar accuracy rates for French and English irregulars. This interplay between language-level and child-level input factors in bilingual acquisition has also been found for French object clitics and Welsh plurals (Paradis, Tremblay, & Crago, 2014; Thomas, Williams, Jones, Davies, & Binks, 2014)

Quantity of input interacts with the quality of that input in terms of how it predicts bilingual children's rate of acquisition. For example, Sorenson-Duncan and Paradis (2017) found that the quantity of L2 input from older siblings was predictive of children's L2 abilities, but the quantity of L2 input from parents was not. This difference was likely due to the older siblings being more fluent L2 speakers than the parents. Both quantity and quality of input are proximal input factors that are influenced by distal factors, like maternal education. For monolinguals, the influence of maternal education on proximal input is that mothers with higher levels of education speak more to their children and provide more complex input (e.g., Huttenlocher, Waterfall, Vasilyeva, Vevea, & Hedges, 2010). The relationship between maternal education and proximal input can be more complex for bilinguals. Sorenson Duncan and Paradis (2018) found that language of education made a difference in how maternal education impacted proximal input factors in bilingual families. For mothers who were educated in the children's L1, higher levels of education were associated with greater use of the L1 with their children; whereas, for mothers who were educated in the L2, higher levels of education meant greater use of the L2. The impact of relative L1 use at home on children's L1 acquisition was stronger than the impact of relative L2 use at home on children's L2 acquisition, due in part to the intervening role of maternal L2 fluency. Both these studies suggest that measuring sheer quantity of input in each language at home might not be sufficient to capture bilingual children's uptake from the input in their home environment (see also Hoff, Welsh, Place, & Ribot, 2014).

While age of acquisition is not strictly speaking an input factor, it is a child-level factor that can change the uptake from the language-level input. Children who begin to learn English around age 4 acquire complex sentences much faster, i.e., with less exposure to input, than younger children acquiring English as their L1 (Paradis, Rusk, Sorenson Duncan, & Govindarajan, 2017). Among children aged 4 to 7, older age of L2 acquisition predicts larger vocabularies and more accurate use of morphology, with amount of L2 input partialled out (Paradis, 2011). It seems that older children can do "more with less" when it comes to the L2 input, likely due to their enhanced cognitive abilities for learning, which could increase uptake. Yang also notes this as a factor in discrepancies between child L1 and adult L2 acquisition.

In sum, bilingual children's access to, and uptake of, language-level input factors is modulated by layers of input variation at the child level. I would be interested in seeing how Yang's formalist Variational Model could account for the greater complexity in the input-acquisition relationship evident in bilinguals.

References

Bybee, J. (2001). *Phonology and language use*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511612886

Bybee, J. (2008). Usage-based grammar and second language acquisition. In P. Robinson & N. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 216–36). New York: Routledge.

Grosjean, F. (2010). *Bilingual: Life and reality*. Cambridge, MA: Harvard University Press. https://doi.org/10.4159/9780674056459

Hoff, E., Welsh, S., Place, S., & Ribot, K. (2014). Properties of dual language input that shape bilingual development and properties of environments that shape dual language input. In T. Grüter & J. Paradis (Eds.), *Input and experience in bilingual development* (pp. 120–140). Amsterdam: John Benjamins.

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of variability in children's language growth. *Cognitive Psychology*, 61, 343–365. https://doi.org/10.1016/j.cogpsych.2010.08.002

Paradis, J. (2011). Individual differences in child English second language acquisition: Comparing child-internal and child-external factors. *Linguistic Approaches to Bilingualism*, 1, 213–237. https://doi.org/10.1075/lab.1.3.01par

Paradis, J., Nicoladis, E., Crago, M., & Genesee, F. (2011). Bilingual children's acquisition of the past tense: A usage-based approach. *Journal of Child Language*, 37, 1–25.

Paradis, J., Rusk, B., Sorenson Duncan, T., & Govindarajan, K. (2017). Children's second language acquisition of English complex syntax: The role of age, input and cognitive factors. *Annual Review of Applied Linguistics*, 37, 1–20. https://doi.org/10.1017/S0267190517000022

Paradis, J., Tremblay, A., & Crago, M. (2014). French-English bilingual children's sensitivity to child-level and language-level input factors in morphosyntactic acquisition. In T. Grüter and J. Paradis (Eds.), *Input and experience in bilingual development* (pp. 161–180). Amsterdam: John Benjamins.

Sorenson-Duncan, T. & Paradis, J. (2017). Look who's talking: effects of sibling versus maternal input in child L2 acquisition. *42nd Boston University Conference on Language Development*, Boston, MA.

Sorenson Duncan, T. & Paradis, J. How does maternal education influence the linguistic environment supporting bilingual language development in child L2 learners of English? *International Journal of Bilingualism*. https://doi.org/10.1177/1367006918768366. Published online by SAGE Journals June 13, 2018.

Thomas, E., Williams, N., Jones, L. A., Davies, S., & Binks, H. (2014). Acquiring complex structures under minority language conditions: Bilingual acquisition of plural morphology in Welsh. *Bilingualism: Language and Cognition*, 17, 478–494. https://doi.org/10.1017/S1366728913000497

Yang, C. (2018). A formalist perspective on language acquisition. *Linguistic Approaches to Bilingualism*, 8(6), 665–706.

Address for correspondence

Johanne Paradis
University of Alberta
Department of Linguistics
4-57 Assiniboia Hall
Edmonton, AB, T6G 2E7
Canada

jparadis@ualberta.ca

Publication history

Date received: 26 June 2018
Date accepted: 15 August 2018

Grammar acquisition and grammar choice in the variationist model

Tom Roeper
University of Massachusetts Amherst

Yang's variable approach to rule-exceptions, sensitive to changing frequency patterns, is a major step forward in capturing the developmental trajectory of acquisition. It provides a new level of granularity in the projection of micro-steps on the acquisition path constrained by UG.

Yang argues exceptional subdomains can reflect opposite parametric choices (see Chomsky, 1986). This leads to Multiple Grammars in everyone's language (e.g., Amaral & Roeper, 2014; Bauke, 2015; Roeper, 2016; Yang, 2016). But grammars must be essentially acquired to be compared: the winner becomes *productive*; the loser is a lexical or constructional exception. In this commentary, I suggest that two refinements can help: (1) A distinction between grammar *acquisition* and grammar *choice*. (2) Articulating a role for semantics and pragmatics in the formulae.

I will show that Pro-drop, Tense, -ed and V2 all show Multiple Grammars during acquisition. Frequency-sensitivity gradually shifts where the productive/exceptional line should be drawn, implying an intricate set of not-yet-visible steps at the microparametric level.

1. Pro-drop: Missing subjects and missing objects

Recognition of *there*-insertion should trigger obligatory subjects, but it is not a logical necessity, since the child gets plenty of *opposite* evidence that subjects are optional (1), so the possibility that pro-drop might be completely productive also has support (see Hyams, 2011; Rizzi, 2005; Hyams et al., 2015 for discussion of pragmatic factors).

(1) a. "seems nice"
    b. "looks good"

https://doi.org/10.1075/lab.18087.roe
Linguistic Approaches to Bilingualism 8:6 (2018), **758–763**.
ISSN 1879–9264 | E-ISSN 1879–9272

c. "going well, isn't it"
d. "did your best, huh?"
e. "gets better and better"

Yet the obligatory-subject grammar is *acquired* before 2 years, as there-insertion indicates (2).

(2) a. Jun (1.7) "there choo-choo train"
b. Eve (1.11) (staring out window:
c. "there no squirrels" (1.11)
d. "there no more these" (2.2)
e. (about tapioca) "there be no more",
f. Peter (2.3) "there no rope here",
g. Naomi (2.11) "there's not enough room". (see Roeper, 2013)

I suggest acquisition occurs when a Strict Interface (SI) is projected through a transparent relation between *syntax, semantics, and pragmatics*. SI separates expletive *there* from locative or early forms of *Presentational there* ("there is a hat!") via input like "there are *no* toys *here*". Such an SI combination creates in a 'triggering *experience*' including pragmatics (pc), with a few clear examples. Only once the obligatory subject grammar is acquired, does the *total* number of *there*-insertion cases become critical for language *choice*, rejecting pro-drop.

The same logic holds for object-dropping. Consider non-adult Adam02 (2.5yrs) "take off" in (3), which is also attested earlier and in many other elliptical cases of non-adult missing objects.

(3) "take off [?]"
"take off here"
"I go take off?"
"I go Adam take off?"
"take out"
"I take"

Perez et al. (2018) propose UG Default Transitivity to explain missing objects. Spanish and English children (5–6yr olds) insert object topics in contexts like (4).

(4) [picture: mom is cooking eggs]
"Look Johnny has as fish!"
"Is Mom cooking?"
English => yes (cooking eggs)
Spanish => no [referring to Mother not cooking fish]

Thirty percent of (4–5yr old English children make the assumption that the Discourse Topic (FISH) fills the object and say "no", as do L2 learners.

Again, substantial parental counter-evidence creates exception classes ("time to pick up, clean up, wash up", less likely *"we threw out") as well as hundreds of nominalizations (takeout, cookout, knockout, strikeout work-around, write-in, walk-up).

Yang's model should therefore include *macro-parameters* (Roberts, 2012) that co-occur with language specific parameters, e,g., Huang's (1981) hot/cool parameter, according to which context can induce empty categories. Yang's model should be enriched to count missing objects twice.

If both macro- and micro-parameters are being set at the same time, then Yang's (2018) observation is predictable: "during no stage of acquisition does the distribution of English-learning children's argument use resemble that of speakers or learners of pro-drop and topic-drop languages…" (p. 671). If indeed children stop omitting subjects and objects at the same time, it would be striking evidence for the macro-parameter.

2. Articles and semantic/pragmatic variation

Yang suggests that a minimal distinction between definite/indefinite articles will drive the comparison sets for *article* acquisition while subtler semantic features come later. I thought children would block subtle *control* by articles until 6–9yrs (Roeper, 1981). Consider (5).

(5) a. Show the bear jumping.
    b. Show the bear *the* jumping.

In (5a), the bear jumps and in (b) someone else does (an arbitrary subject). The article converts V to N and blocks control of PRO in the VP [PRO jumping]. To my surprise, many three-year-olds knew this. This experiment demonstrated that sophisticated UG properties could be instantly projected from surely rare data, i.e., non-control gerunds.

In contrast, Maximality with definite articles arises quite late (*John ate the cake* = whole cake). Why? Maximality is pragmatic, not directly represented in syntax, which raises the question: Which semantic variants need to be represented in possible grammar comparisons at each step? We must allow both languages with no articles and those with articles (see Cherchia's 1998, Nominal Mapping Parameter). In some languages, articles are obligatory: but there are also distinctions of part-whole relations (cake with cream on *the* top) inalienable possession (*hit him in the eye*), etc. The choicepoints will be many across languages; for instance, Kupisch (2006) observes 14 different discernible factors (like speaker awareness, specificity, visibility in context, etc.) for articles. Hence, articles further illustrate the need for micro-parameters and the pragmatic influences of Multiple Grammars and statistical calculation on acquisition. This is not easy to do empirically, but it is better to labor to discover how to incorporate pragmatics than to leap to a misleading higher level of abstraction that is exclusively syntactic.

3. Tense and semantic variation

English separates present generic from progressive readings: *John reads =/=John is reading* However, L2 German speakers of English often confuse the two. To get English versus German straight, they must be separate grammars for both L1 and L2. Nevertheless, English does not distinguish past and past progressive*: John read a book during dinner/ John was reading a book during dinner* are both imperfect. *I read (past) the book* entails completion.

If the child assumes that *read* is past tense only, just like *-s* is present not progressive, then he should assume that there is a competition among those readings and project two grammars. If he does not for *-ed*, then he should not for *-s* either. Examination of early past tenses might answer the question: Do the early *-ed* and/or *-s* cases cover progressive meaning or are they all excluded at first? Refined semantic variations must be explicit in modeling the acquisition path, not just *past*. Semantic variation may make grammar choice more sensitive to small frequency decisions.

4. V2

V2 is another major parameter allowed in both German and English. Children hear plenty of cases like "*here comes Daddy*" or "'*No' said Mama*" in stories. The verb *be* supports V2 as well (*John is not here* (but not: **John plays not here*). If statistical sensitivity is to tokens, not only types, thousands of forms like *"John isn't big*" support V2 in English for a child. How does she eventually make it non-productive? Wexler (2011) argues that children get V2 early in German, and Yang says V2 is late because Obj-verb-Subj (*Fleisch isst Hanns* (*meat eats John*)) is infrequent. However, the full array of XPs in initial position is realized even more slowly in German (Roeper, 1973), in particular for fronted infinitives [Fussball zu spielen macht Spass (to play football makes fun). When there are enough types, a list is converted into just XP, and V2 is *chosen* as productive in German, though it is *acquired* earlier. English never gets enough, so V2 is unproductive.

In sum, Yang's approach, augmented by an acquisition/choice distinction and subtle semantics should reveal a frequency-sensitive profile to both micro- and macro-parameters.

References

Amaral, L., & T. Roeper. (2014). Multiple grammars and second language representation. *Second Language Research*, 30(1), 3–36. https://doi.org/10.1177/0267658313519017

Bauker, L. (2015). *Content matching in idiom and compounds. A comparative analysis*. Paper presented at BCGL8 Grammar and Idioms, Brussels.

Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin, and Use*. New York: Praeger.

Hyams, N. (2011). Missing Subjects in Early child Language, in J. De Villiers and T. Roeper (Eds.), *Handbook of Generative Approaches to Language Acquisition*, Dordrecht: Springer.

Hyams, N., Mateu, V., Orfitelli, R., Putnam, M., Rothman, J., & Sánchez, L. (2015). Parameters in language acquisition and language contact. In Fábregas, A., Mateu, J., & Putnam, M. (Eds.), (2015). *Contemporary Linguistic Parameters*. New York: Bloomsbury Publishing.

Kupisch, T. (2006). The acquisition of determiners in bilingual German-Italian and German-French children (Doctoral dissertation). Hamburg University, Hamburg.

Perez, A., M. Pirvulescu, & Y. Roberge. (2018). *Direct objects and language acquisition*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781139086264

Rizzi, L. (2005). Grammatically-based target-inconsistencies in child language. In Deen, K., Nomujra, J., Schultz, B. and Schwartz, B. D. (Eds.), *Proceedings of the inaugural conference on generative accproaches to language acquisition – North America (GALANA)*. Cambridge, MA, UCONN / MIT Working Papers in Linguistics.

Roberts, I. (2012). Macroparameters and minimalism: A programme for comparative research. In C. Galves, S. Cuyrinom, R. Lopes, F. Sandolo, & J. Avelar (Eds.), *Parameter theory and linguistic change* (pp. 320–369). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199659203.003.0017

Roeper, T. (1981). On the deductive model and the acquisition of productive morphology. In C. L. Baker, & J. McCarthy (Eds.), *The logical problem of language acquisition* (pp. 129–150). Cambridge, MA: MIT Press.

Roeper, T. (2013). *There*-insertion: How internal Merge guides the acquisition path. In M. Becker, J. Rothman, & J. Grinstead (Eds.), *Generative linguistics and acquisition: Studies in honor of Nina M. Hyams* (pp. 249–270). Amsterdam: John Benjamins Publishing. https://doi.org/10.1075/lald.54.10roe

Roeper, T. (2016). Multiple grammars and the logic of learnability in second language acquisition. *Frontiers of Psychology*, 7(14).

Schafer, R. & Roeper, T. (2000). *Cross Level Triggers: How an Expletive Triggers the Acquisition of a Discourse Anaphor*. Ms., University of Massachusetts, Amherst.

Yang, C. (2016). *The price of linguistic productivity: How children learn to break rules of language*. Cambridge, MA: MIT Press.

Yang, C. (2018). A formalist perspective on language acquisition. *Linguistic Approaches to Bilingualism*, 8(6), 665–706.

Address for correspondence

Tom Roeper
Linguistics Department (Integrative Learning Center)
University of Massachusetts Amherst
650 North Pleasant Street
Amherst, MA 01003
USA

roeper@linguist.umass.edu

Publication history

Date received: 22 September 2018
Date accepted: 22 September 2018

Towards eliminating arbitrary stipulations related to parameters

Linguistic innateness and the variational model

Jason Rothman[1,2,3] and Noam Chomsky[4,5]
[1]University of Reading / [2]UiT the Arctic University of Norway / [3]Facultad de Lenguas y Educación, Universidad Nebrija. Madrid, Spain / [4]Massachusetts Institute of Technology / [5]University of Arizona

The Variational Model (VM) (Yang, 2002, 2016) is inspired by evolutionary models of biological change; in a way, a grammar survives if indeed it is the fittest; "A grammar whose competitor is penalized more often will be learned faster" (Yang, 2018, p. 668). According to the VM, the developing learner simultaneously entertains a number of possible grammars associated with stochastic weighting. Underlying grammatical representations (intermediary and ultimately attained ones alike) are contingent on the child's tracking of probabilistic distributions in response to available input. In our view, Yang's approach is an exciting step forward for understanding the dynamics of language acquisition, not the least because it reconciles some seemingly contentious debates. Highlighting the indispensability of input for determining the ultimate system, it carves out a proper place for innate determination of the set of potential grammars and linguistic categories, along with a learning theory that may well be in part human-specific.

In the present paper, Yang (2018) describes and explains the mathematics behind the VM and its coverage of impressive data sets in child (monolingual) acquisition. The most novel part of the paper involves the application of the VM to the domain of adult second language acquisition, uncovering at the same time potential insights that could eventually be extended/modified to bilingualism more generally (e.g., heritage language bilingualism, first language (L1) attrition, simultaneous child bilingualism). Given the complexities and range of topics touched upon, there is no shortage of points worthy of thoughtful commentary. We expect that several colleagues will focus on the relative potential for an (increasingly) successful application of Yang's model to bilingualism. And so, we will focus on something else that should be of ubiquitous concern to all who study language

https://doi.org/10.1075/lab.18075.rot
Linguistic Approaches to Bilingualism 8:6 (2018), **764–769**.
ISSN 1879-9264 | E-ISSN 1879-9272

acquisition/attrition, be it monolingual, bilingual or multilingual in childhood or adulthood. We will focus on the following two interrelated topics: understanding (a) the Variational Model as a natural development of generative models of language acquisition and (b) how arbitrary stipulations of parameter values can be eliminated in terms of the learning theory he has developed.

The recognition that input plays an indispensable role for mental grammatical construction is shared by all cognitive based approaches to language acquisition, as is a (certain level of) innateness. Usage-based theories are fully innatist; the difference from a generative grammar view is the claim that what is innate is strictly domain-general. The investigation of generative grammar, when concerned with acquisition, has led to the conclusion that elements of domain-specificity have a crucial role in determining the hypothesis space (the set of potential grammars). In this context, it is useful to keep in mind that generative grammar is the theory of the possible states attained. As such, generative grammar itself is neutral about how the final state (a grammar/I-language, or probability distribution over them) is attained. It is beyond question that much of the resulting idiosyncratic properties of particular grammars, including the specificity/size of vocabulary and how we use language, depend on linguistic exposure.[1] There is no principle of generative grammar about that, instead generative grammar is about the phenotype. It is, then, an empirical conclusion that the innate basis for language includes domain-specific principles (Universal Grammar, UG).

There has been substantial progress in determining what the innate principles might be, though much remains unknown. Nevertheless, they do appear to be unique to humans and crucial, even for word learning. Let's take, for example, the very simple principle that yields structure dependence, a core feature of language as distinct from any symbolic system known in the animal world, all of which seem to use the very simple properties of linear order (as other human cognitive processes do). The empirical question is to determine the principles that enter into language acquisition, and to determine which are language-specific. Yang's work is one of the rare contributions identifying what may be a more general learning principle, which relates to UG, specifically showing that at least some valuation of parametric variation may not have to be stipulated for particular languages. In doing so, it

1. Alignment with linguistic exposure is of course not a simplistic one-to-one mapping, even for word learning or aspects thereof (see Gleitman & Fisher, 2005; Lidz, Gleitman & Gleitman, 2003). For example, the complex meaning of even the simplest words, those used to refer, is acquired on very few exposures. Exposure seems to involve mostly Saussurean arbitrariness. In fact, there is no imaginable course of experience that could yield these meanings, revealed when one looks at them carefully.

leaves the UG framework (right or wrong) otherwise unchanged, eliminating previous stipulations proposed to account for descriptive facts/observations.

What better distinguishes generative approaches and usage-based/emergentist ones has never rested on whether or not the language acquisition device is context and/or frequency sensitive – it must be – but on whether some aspects of language are domain-specific, or none at all. Chomsky (1959) is often cited as the genesis of the generative linguistic enterprise, and so it is worth revisiting statements as they originally appeared:

> As far as acquisition of language is concerned, it seems clear that reinforcement, casual observation, and natural inquisitiveness (coupled with a strong tendency to imitate) are important factors, as is the remarkable capacity of the child to generalize, hypothesize, and "process information" in a variety of very special and apparently highly complex ways which we cannot yet describe or begin to understand, and which may be largely innate, or may develop through some sort of learning or through maturation of the nervous system. The manner in which such factors operate and interact in language acquisition is completely unknown. It is clear that what is necessary in such a case is research, not dogmatic and perfectly arbitrary claims, based on analogies to that small part of the experimental literature in which one happens to be interested. (Chomsky, 1959, p. 43)

In 2018, almost 60 years on, several parts of the above quote could be updated although the essence remains. It is as undeniable today as it was in 1959 that "*[…] casual observation, and natural inquisitiveness (coupled with a strong tendency to imitate) are important factors*". We do know much more today about how factors mentioned above "*operate and interact in language acquisition*", although few would deny we have much more to uncover. Over the last six decades, various proposals have articulated opposing views on what underlies *"the remarkable capacity of the child to generalize, hypothesize, and "process information" in a variety of very special and apparently highly complex ways"* although we are now better equipped to describe and eventually understand them. The question of what subset of this capacity is "*largely innate or may develop through some sort of learning or through maturation of the nervous system*" is still an open one, although nearly sixty years of theorizing within generative grammar have narrowed the gap, reducing many specific assumptions about more general principles of UG and computational efficiency.

As Yang points out, the theory of innate parameters – the idea that UG provides a set of highly specific universal principles with constrained options to be set by the child – has been developed extensively since it was proposed 40 years ago. There has been substantial work on the nature and organization of parameters, on ways to reduce the search space in acquisition, and in recent years, on ways that parameters might be reduced to principles of efficient computation (e.g., Berwick & Chomsky, 2016; Epstein, Obata, & Seely, 2017). Parameters in the classic sense were (and remain) a fundamental construct. At the conceptual level, the theory offered the first suggestion about a feasible acquisition system, with bounded choices. It led to inquiry into languages of scale and depth that goes beyond anything in the history of the subject, along with very fruitful work on the nature and structure of linguistic variation. Its legacy includes fine-grained descriptions of grammatical domains, a program designed to describe and explain linguistic variation and observations of the co-occurrence of grammatical phenomena that otherwise would not be expected (see e.g., Fábregas, Mateu, & Putnam, 2015; Karimi & Piattelli-Palmarini, 2017 *inter alia*). It led to new work in diachronic linguistics, previously unformulable, about setting of parameters and how it changed (Lightfoot & Westergaard, 2007). It injected innovation into psycholinguistics, both processing and acquisition (see Hyams et al., 2015; Sanz, Laka, & Tanenhaus, 2013). The granularity of the theory lent itself well to making order-of-acquisition predictions, explaining patterns of production, providing a toolkit to describe developing language in formal terms, as well as offering a seemingly good-enough answer to the logical problem of acquisition. Related to the Chomsky (1959) quote, it was a proposal on what "may be largely innate".

Yang's work doesn't bear on the very existence of parameters but rather on how they are valued. There are three questions to face about parameters: (1) what they are: the options available for language (2) how they are set in particular languages and (3) where they come from.[2] Yang's work deals with (2), essentially assuming (1). Let's consider the work on double objects. He assumes that there are two options: double object/(dative) to-phrase, which constitutes the parameter. Innumerable other ways of organizing the (limited) data available to the child are ignored. The Tolerance Principle very insightfully shows how the setting can be determined (caused possession, etc.). This eliminates the unwelcome need to stipulate values of parameters. But it doesn't eliminate the parameter per se, which

2. Given that it is hard to imagine parameters evolved, serious work regarding (3) is in progress. Take the head parameter. If nature selected the simplest way to yield discrete infinity (Merge), then the structures generated internally, and semantically/pragmatically interpreted, have no linear order. But the articulatory system, which is independent of language, requires linear order. So there is a mismatch between two unrelated systems (it's a bit different with sign, which has more dimensions available). Externalization must assign linear order, and there are two ways, head-first/-last. But there is no parameter. Samuel Epstein and his group have developed such ideas further, discussing cases of variation where, they argue, it is a matter of a choice that has to be made in rule ordering where UG leaves it open. Such approaches actually do get rid of parameters – but what lies ahead is an immense task, posed clearly by the fundamental work of Mark Baker, Giuseppe Longobardi, Richard Kayne, and many others.

determines the options, excluding innumerable others. Here we see the greatest contribution of Yang's model: its overcoming the stipulation of values for innate linguistic parameters. In doing so, Yang's work focuses on gradual yet highly constrained development. His algorithmic approach provides specific answers to the question when input data yield rule formation rather than listing, with quite remarkable results.

Yang's work shows very effectively how unlearned/unlearnable properties of language interact with specific linguistic experience in child language acquisition, a very significant contribution. With Yang, we hope that the VM and specifically the Tolerance Principle will enable further progress in understanding these interactions in the many domains that are yet to be explored. Some 20 years on, it is still true today that:

> The field is changing rapidly under the impact of new empirical materials and theoretical ideas. What looks reasonable today is likely to take a different form tomorrow […] Though the general framework remains, the modifications at each point are substantial […] The end result is a picture of language that differs considerably from its immediate precursors. Whether these steps are on the right track or not, of course, only time will tell. (Chomsky, 1995, p. 9)

References

Berwick, R. C., & Chomsky, N. (2016). *Why only us: Language and evolution*. MA: MIT Press.

Chomsky, N. (1959). Verbal behavior (a review of Skinner's book). *Language*, 35, 26–58. https://doi.org/10.2307/411334

Chomsky, N. (1995). *The minimalist program*. MA: MIT Press.

Epstein, S. D., Obata, M., & Seely, T. D. (2017). Is linguistic variation entirely linguistic? *Linguistic Analysis*, 41(3), 481–516.

Fábregas, A., Mateu, J., & Putnam, M. (Eds.), (2015). *Contemporary linguistic parameters*. NY: Bloomsbury Publishing.

Gleitman, L., & Fisher, C. (2005). Universal aspects of word learning. In McGilvray, J. (Ed.), *The Cambridge companion to Chomsky*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CCOL0521780136.007

Hyams, N., Mateu, V., Orfitelli, R., Putnam, M., Rothman, J., & Sánchez, L. (2015). Parameters in language acquisition and language contact. In Fábregas, A., Mateu, J., & Putnam, M. (Eds.), (2015). *Contemporary linguistic parameters*. NY: Bloomsbury Publishing.

Karimi, S. & M. Piattelli-Palmarini (Eds.) (2017). Parameters: What are they? Where are they? Special issue, *Linguistic Analysis*, 41(3 & 4).

Lidz, J., Gleitman, H., & Gleitman, L. (2003). Understanding how input matters: verb learning and the footprint of universal grammar. *Cognition*, 87(3), 151–178. https://doi.org/10.1016/S0010-0277(02)00230-5

Lightfoot, D., & Westergaard, M. (2007). Language acquisition and language change: Interrelationships. *Language and Linguistics Compass*, 1(5), 396–415. https://doi.org/10.1111/j.1749-818X.2007.00023.x

Sanz, M., Laka, I., & Tanenhaus, M. K. (Eds.). (2013). *Language down the garden path: The cognitive and biological basis for linguistic structures*. OUP Oxford. https://doi.org/10.1093/acprof:oso/9780199677139.001.0001

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

Yang, C. (2016). *The price of linguistic productivity: How children learn to break the rules of language*. MA: MIT Press.

Yang, C. (2018). A formalist perspective on language acquisition. *Linguistic Approaches to Bilingualism*, 8(6), 665–706.

Address for correspondence

Jason Rothman
University of Reading
School of Psychology & Clinical Language Sciences
Harry Pitt Building
Earley Gate, Reading, RG6 7BE
United Kingdom

j.rothman@reading.ac.uk

Co-author details

Noam Chomsky
University of Arizona
Department of Linguistics
P.O. Box 210025
Tucson AZ 85721
USA

chomsky@mit.edu

Publication history

Date received: 30 July 2018
Date accepted: 15 August 2018

The principles of scientific inquiry

Caroline F. Rowland
Language Development Department, Max Planck Institute for Psycholinguistics, The Netherlands / ESRC LuCiD Centre, University of Liverpool

Cognitive science is a multi-disciplinary endeavour. In many ways, this is an advantage; the strongest theories of L1 and L2 acquisition are supported by a consensus of evidence across disciplines such as psychology, linguistics, neuroscience and computational science (Monaghan & Rowland, 2017). However, multi-disciplinarity can also lead to confusion because different disciplines approach similar research questions from diverse perspectives, and bring with them very different, often unspoken, assumptions.

Given this, an important factor that unites all the cognitive sciences is adherence to the principles of Popperian deductive science (Popper, 1959). For Popper, as for most modern scientists, the criterion distinguishing science from non-science is falsifiability. A theory is only scientific if it makes predictions that are potentially incompatible with empirical observations. Theories that are compatible with all observations, either because they are endlessly modified to accommodate such observations, or because they are so vague as to be consistent with all possible observations, are unscientific. For many cognitive scientists, there is something frustrating about a theory that makes no falsifiable predictions, a frustration shared by Popper with respect to the psychoanalytic theories of Freud and Adler, which "do not exclude any physically possible human behaviour" (Popper, 1974, p. 985).

Thus, I am pleased that Yang (2018) shows an admirable commitment to all four principles of Popperian scientific reasoning in this article. First, his models adhere to the principle that theories have to be internally consistent; they cannot contain any mutually incompatible statements (i.e. contradictions). For example, the Variational model (Yang, 2002) assumes that (a) the child entertains a number of grammars in the learning space, (b) these grammars are associated with probabilities or weights, and (c) learning takes place as the child makes changes to the probabilistic distribution of the grammars in response to the input. The

https://doi.org/10.1075/lab.18056.row
Linguistic Approaches to Bilingualism 8:6 (2018), 770–775.
ISSN 1879-9264 | E-ISSN 1879-9272

consequent prediction is, thus, that (d) "the amount of unambiguous evidence for parameter values in child-directed input corpora [will] correlate with the developmental time course of the parameters" (p. 669). All of these statements are logically compatible.

The second principle requires that the logical form of the theory must be explicit and formulated in such a way that its basic statements correspond to experience. In other words, the theory must describe a possible real world. Here, there are some areas in which Yang's assumptions seem underspecified (e.g. how the Variational model solves Pinker's (1984) bootstrapping problem), but overall much of what Yang proposes is explicit enough to be testable against experience. In particular, his mathematical models describe plausible learning mechanisms – mechanisms that are possible given what we already know about learning in the brain. For example, the Variational model incorporates a learning parameter that captures the fact that some children are slower to process linguistic information than others (Fernald, Perfors, & Marchman, 2006).

The third principle states that it must be possible to compare, favourably, the new theory to existing ones. Again, Yang's theories adheres to this principle. For example, he shows that the Variational model provides a better explanation of the null subject stage in English acquisition, in which children omit obligatory subjects (e.g. _ *want cookies*), than traditional parameter setting theories. These theories predict, wrongly, that children's language will resemble that of speakers of pro- or topic-drop languages during this stage. The Variational model, however, is compatible with the pattern of subject omission in children's speech.

Fourth and finally, a scientific theory must be testable "by the empirical application of the conclusions derived from it" (Thornton, 2017: Section 4). It must make predictions about the behaviour of the world that can be tested against observations in the world. If the predictions are supported by the evidence, the theory is corroborated. If not, the theory cannot be completely correct. Again, Yang adheres to this principle (at least in part, see below) because his predictions are clearly stated and tested against empirical evidence.

For example, in Section 3.2, Yang (2018) summarises work from 2013, in which he tested the predictions of two theories of determiner acquisition. The early productivity theory proposes that children's "combinatorial productivity [in their use of determiners] is on full display from the earliest testable stage of multi-word combinations" (2018, p. 677). The logical prediction of this hypothesis is that young children's productivity with determiners, defined as their ability to combine determiners *a* and *the* freely with nouns, will be as productive as that of adults. For example, if 19% of the nouns used by adults are paired with both *a* and *the* in a speech sample, then approximately 19% of nouns used by children should occur with both *a* and *the* in a similar size sample (combinatorial productivity).

The lexical specificity theory, however, claims that children start out with lexically specific knowledge of how to combine *a* and *the* with certain nouns (often in semi-formulaic phrases such as *where's the X, that's a Y)*, and only later develop the understanding that determiners can combine with nouns productively (Yang cites Pine & Lieven, 1997, but see also Pine, Freudenthal, Krajewski, & Gobet, 2013; Pine & Martindale, 1996). The logical prediction of this hypothesis is that young children's determiner use will be more restricted than that of adults (e.g. if 19% of an adult's nouns occur in a sample with both determiners, significantly fewer of the children's nouns will occur with both determiners).

Testing these two hypotheses is made difficult by the fact that both adult and child speech follows a Zipfian distribution, in which a small number of words occur very often, but most words are rare even in large corpora. This means that even adults may show limited combinatorial productivity, because the chances of these low frequency nouns occurring with both *a* and *the* in a given speech sample are small. Using data from corpora of children's and child directed speech, Yang (2013) showed that there is no significant difference in the productivity of children's and adult's determiner usage, once we control for the expected distribution of nouns and determiners. This evidence supports the early productivity theory's prediction that young children's use of determiners *a* and *the* is adult like, and not the prediction of the lexical specificity theory. Interestingly, the prediction does not hold for adult L2 learners of English, who show more limited combinatorial productivity (see Section 5.2, Table 4, p. 691). Thus, in accordance with Popper's fourth principle, Yang concludes that the extension of the theory to L2 acquisition does not hold, at least in the theory's current form.

However, with regards to this fourth principle of falsifiability, I would urge Yang to go further, and directly address evidence which other authors have provided on this question. In the spirit of Popperian scientific enquiry, theories need to be evaluated against all the evidence against them. There are, thus, some surprising omissions in this article.

In particular, I was surprised there was no discussion of Pine et al.'s (2013) detailed work on determiner acquisition, which built on Yang's (2013) own careful mathematical analysis. Pine et al. agreed with Yang that the expected productivity scores should take into account the Zipfian distribution of language. However, they also argued that Yang's analysis underestimated the productivity of adult speech because it did not control for the fact that adults know more nouns than children: "it is not based on a controlled set of nouns, but on an adult corpus that includes between 5 and 16 times as many different nouns as any of the child corpora being analysed" (Pine et al., 2013, p. 349). Given the Zipfian distribution of nouns and determiners in speech, this means that "the proportion of nouns that occur with low frequency [and are thus unlikely to show overlap] in the adult corpus is likely to be considerably higher than the proportion of nouns that occur with low frequency in the children's corpora" (Pine et al., 2013, p. 349). Thus, they argued, Yang's analyses substantially underestimates the productivity of adult speech, invalidating the comparison with child speech.

In a detailed set of analyses, Pine et al. (2013) showed that failing to control for noun identity does, indeed, depress adult productivity scores (study 1). They then compared child and adult speech, restricting the analysis to the nouns that both adults and children produced, thus removing the confound. Contrary to Yang's (2013) conclusions, children's determiner use was significantly less productive than that of adults (study 2). There was also a developmental cline in productivity, with children's determiner use becoming more productive with age (study 3), again contrary to the prediction of the early productivity account.

Of course, this is only a single paper (though note that Meylan, Frank, Roy, and Levy (2017) replicate the findings of Pine et al., study 3). Popper himself acknowledged that a single counter-example is not sufficient to falsify a theory in practice. Non-corroboration is not necessarily falsification because no one observation is free from error (see also Lakatos, 1978). Thus, we should not dismiss a theory on the basis of one study alone.

However, Pine et al.'s (2013) evidence is not discussed at all in the keynote article, bar a passing reference in a footnote. I do not think Popper would approve. Even if we disagree with the conclusions of our critics, the solution is to engage with their evidence, both on an intellectual level using logical argumentation, and on an empirical level in replication attempts. Yang's three mathematical models have extremely interesting implications for research on both L1 and L2 acquisition. However, we must be careful to weigh up all the evidence when we evaluate our models. "Those among us who are unwilling to expose their ideas to the hazard of refutation do not take part in the scientific game" (Popper, 1959, p. 280).

Funding

This work was supported by a grant from the Economic and Social Research Council ES/L008955/1 awarded to Caroline Rowland.

Acknowledgements

The author was supported by the International Centre for Language and Communicative Development at the University of Liverpool (LuCiD).

References

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, 42(1), 98–116. https://doi.org/10.1037/0012-1649.42.1.98

Lakatos, I. (1978). The methodology of scientific research programmes. In J. Worrall & E. Currie (Eds.), *Philosophical papers ; vol. 1* (p. viii, 250). Cambridge, UK: Cambridge University Press.

Meylan, S. C., Frank, M. C., Roy, B. C., & Levy, R. (2017). The emergence of an abstract grammatical category in children's early speech. *Psychological Science*, 28(2), 181–192. https://doi.org/10.1177/0956797616677753

Monaghan, P., & Rowland, C. F. (2017). Combining language corpora with experimental and computational approaches for language acquisition research. *Language Learning*, 67, 14–39. https://doi.org/10.1111/lang.12221

Pine, J. M., & Martindale, H. (1996). Syntactic categories in the speech of young children: The case of the determiner. *Journal of Child Language*, 23(2), 369–95. https://doi.org/10.1017/S0305000900008849

Pine, J. M., & Lieven, E. V. M. (1997). Slot and frame patterns and the development of the determiner category. *Applied Psycholinguistics*, 18(02), 123–138. https://doi.org/10.1017/S0142716400009930

Pine, J. M., Freudenthal, D., Krajewski, G., & Gobet, F. (2013). Do young children have adult-like syntactic categories? Zipf's law and the case of the determiner. *Cognition*, 127(3), 345–360. https://doi.org/10.1016/j.cognition.2013.02.006

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA.: Harvard University Press. https://doi.org/10.1007/s13398-014-0173-7.2

Popper, K. R. (1959). *The logic of scientific discovery*. London: Hutchinson.

Popper, K. R. (1974). *The philosophy of Karl Popper*. (P. A. Schilpp, Ed.). La Salle, Illinois: Open Court Publisher.

Thornton, S. (2017). Karl Popper. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Stanford University. Retrieved from https://plato.stanford.edu/archives/sum2017/entries/popper/

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

Yang, C. D. (2013). Ontogeny and phylogeny of language. *Proceedings of the National Academy of Sciences*, 110(16), 6324–6327. https://doi.org/10.1073/pnas.1216803110

Yang, C. (2018). A formalist perspective on language acquisition. *Linguistic Approaches to Bilingualism* 8(6), 665–706.

Address for correspondence

Caroline F. Rowland
Max Planck Institute For Psycholinguistics
Wundtlaan 1, 6525 XD Nijmegen
The Netherlands

Caroline.Rowland@mpi.nl

Publication history

Date received: 24 June 2018
Date accepted: 15 August 2018

Back to our roots

Roumyana Slabakova
University of Southampton / UiT The Arctic University of Norway

With its focus on the transition from one state of knowledge to another, Charles Yang's (2018) epistemological paper is a welcome addition to the generative perspective on language acquisition. It is an example of how generative scholars have begun to take into account usage-based explanations for some (but crucially not all) language acquisition facts. We are still waiting for usage-based scholars to do the same in the opposite direction. The Variational Learning Model (Yang, 2002) combined input effects with traditional-style universal parameterized knowledge to explain why parameters were acquired by children at different ages. The Tolerance Principle is "a method by which the learner evaluates potentially productive hypotheses about language.… a quantitative measure of the input data is absolutely crucial" (Yang, 2018, p. 694). One is left with the understanding that variational learning and input calculations are both necessary and sufficient to explain language acquisition. While I am very sympathetic to this "mechanistic" approach and I have always considered The Variational Model to be on the right track, in this commentary I am going to challenge the approach by pointing to some second language acquisition findings it (still) cannot explain.

There is no doubt that adult second language acquisition is fraught with a lot of interpersonal variation, both in process as well as outcome. However, there are interesting observations and generalizations to be made pertaining to the process as a whole that a simple input-plus-parameters approach cannot account for. One such observation is based on the Minimalist proposal that functional categories, for example Tense Phrase (TP), host a multitude of grammatical meanings (e.g., past tense, perfective aspect, phi features) that may be related to a single piece of inflectional morphology (*-s, -ed*) or syntactic movement. In this sense, the overt morphology conflates these grammatical meanings because it is a surface reflex of all of them. How and when are they acquired? One way to operationalize this multitude is to postulate that the grammatical meanings correspond to features, and to look more closely at the differential acquisition of features depending in the complexity of their conditioning environments, as Lardiere (2009) has done.

https://doi.org/10.1075/lab.18062.sla
Linguistic Approaches to Bilingualism 8:6 (2018), 776–781.
ISSN 1879-9264 | E-ISSN 1879-9272

Another way is to argue (based on the Borer-Chomsky conjecture) that some syntactic and semantic information comes for free, after acquisition of features is effected, as the Bottleneck Hypothesis does (Slabakova, 2008, 2016). An input-based approach would predict that the different TP-related meanings are acquired together, contrary to fact.

I will exemplify this observation with a well-known account from White (2003, p. 187–193). The comparison in Table 1 includes data from Lardiere (1998a, b) and from Li (2012), in order to maintain the native (Chinese) languages constant. Lardiere's subject, Patty, is a Hokkien and Mandarin-bilingual adult learner of English. Li's participants are six Mandarin-native children aged 7 to 9 acquiring English in a naturalistic environment in the USA. Patty's performance is considered to be at end-state, in the sense that it is deemed she will not develop it further. The children's performance is captured longitudinally for eight months, starting when they had been in the USA for four months, so they will clearly continue to develop. The data from Patty and the children, however, show uncanny parallels.

Table 1. Percentage accurate suppliance of grammatical morphemes and the syntactic effects associated with the functional category TP (following White, 2003)

	3sg agreement	**Past tense on lexical verbs**	**Suppletive forms of *be***	**Overt subjects**	**Nom. case on subjects**	**V in VP (no verb raising)**
Lardiere (1998a, b)	**4.5**	**34.5**	90	98	100	100
Li (2012)	**16**	**25.5**	93	100	100	–

What is especially striking in the data presented in Table 1 is the clear dissociation between the accuracy of verbal inflection (ranging between 34.5% and 4.5%) and the various syntactic phenomena related to it, like providing overt subjects, marking nominative case on the subject, and the verb staying in VP (above 98% accuracy). It seems that, even though Patty and the children do not produce the overt morphemes *-s* and *-ed* very often, they still know what the morphemes stand for and what other syntactic processes they regulate in the sentence (see also Yusa, this volume). Another observation is that *-ed* is supplied more often than *-s*, perhaps due to the redundancy of the latter. In view of such comparisons, a legitimate question is: what is being learned until the functional morphology is supplied to criterion? Some features expressed in the same functional category (TP in this case) are acquired much earlier than others. While the Variational Model and the Tolerance Principle successfully use indirect evidence of acquisition (for example, percentages of dropped subjects in adjunct and argument questions for the null subject parameter), an added challenge would be to account for the dissociation of syntactic knowledge and morphological suppliance in L2 production. At least as far as I understand it, the Tolerance Principle would struggle with such data.

Another challenge for the Variational Model and the Tolerance Principle could come from proposals that feature realization depends on feature co-occurrence hierarchies. Not only is the pure number of features that the affix reflects important; it is also important that nodes and their features are arranged by the grammar in feature hierarchies (Harley & Ritter, 2002). One example, provided by Lardiere (2017), comes from 3rd person singular subject pronouns in Patty's production. Many learners of English, including Patty, often confuse the masculine and feminine forms of 3rd person singular pronouns.

> In Patty's case, one might consider that her (spoken) L1 Chinese fails to distinguish pronominal gender and therefore this would be a simple issue of L1 transfer; however, neither does her L1 distinguish pronominal case marking, which in her L2 English was perfect. In other words, while Patty often confused *him* or *his* with *her*, or *she* for *he*, she never confused *he* for *him* or *his*, or *she* for *her*.
> (Lardiere, 2017, p. 56)

These are very specific, feature-related errors that are difficult to explain based on frequency in the input, because the errors do not appear in the input. Lardiere's explanation lies in a feature hierarchy where case (nominative, accusative, genitive) is higher than person/number features, themselves higher than gender (masculine, feminine). She argues that the more deeply embedded a feature is within a feature co-occurrence hierarchy, the less detectable it is and the more difficulty is expected in its retrieval. Again, I believe the Variational Model and the Tolerance Principle as they stand do not offer a clear path to an explanation.

Finally, let us look at successful L2 acquisition where the input appears to be insufficient for acquisition and cannot account for the data. Bruhn de Garavito (2011) discusses knowledge of two structures in Spanish which, on the surface, are expressed by exactly the same string of words, the impersonal passive in (1) and the inchoative in (2). In both sentences the reflexive clitic *se* is obligatory and the verb agrees with the Theme argument. The latter NP displays certain surface properties of subjects, in particular agreement with the verb and the ability to *pro*-drop. In spite of the similarities, these two structures exhibit important interpretive and syntactic differences.[1]

1. In sentences such as (1) there is an implied agent. In addition, the NP also exhibits properties only associated with objects, such as the possibility of subextraction, differential object marking and secondary predication. In contrast, in sentences such as (2), there is no implied agent and the Theme NP only exhibits subject-like properties.

(1) *Se vendieron las faldas.* (Impersonal passive)
se sold-PL the skirt-PL
'The skirts were sold.'

(2) *Se mancharon las faldas.* (Inchoative)
se stained-PL the skirt-PL
'The skirts got stained.'

As Bruhn de Garavito argues, analogy would lead the learner to assume these constructions are syntactically identical. In addition, "[a]ssociation of the string with a particular situation would not be any help given that in every situation the same string is the appropriate response" (Bruhn de Garavito, 2011, p. 112). However, as she demonstrated, early (before 6 years) and late (after puberty) bilinguals were able to acquire the syntactic distinctions between these two constructions successfully. There were no differences between the judgments of both learner groups on this contrast, where the knowledge could not be derived from input alone. The author argued that the input under-determined actual knowledge. An approach that focuses solely on input-based learning algorithms would be hard put to provide an explanation.

Of course, cases like these, and many others, constitute the hallmark argument for the existence of UG that generative SLA has been proposing since the 1980s, (see White, 1989, 2003 for review). The Variational Model and the Tolerance Principle provide an excellent explanation for how acquisition likely happens in the cases where the input and universal linguistic knowledge likely conspire. As Rothman and Slabakova (2018) argued, there are also many areas of language acquisition where universal knowledge is not needed, hence usage-based and generative approaches make the same predictions. But we cannot ignore the many cases like the ones pointed out in this commentary where input plus parameter settings are not sufficient and learner knowledge is both under-determined by the linguistic input and under-represented by learner production. Generative L2 researchers are waiting for acquisition theory to take forward these relevant acquisition findings into future development of acquisition modeling.

Acknowledgements

I am grateful to the LAB epistemological issue editors for inviting me to comment on the paper and to Laura Dominguez for feedback on an earlier version of this commentary.

References

Bruhn de Garavito, J. (2011). Subject/object asymmetries in the grammar of bilingual and monolingual Spanish speakers: Evidence against connectionism. *Linguistic Approaches to Bilingualism*, 1(2), 111–148. https://doi.org/10.1075/lab.1.2.01bru

Harley, H., & Ritter, E. (2002). Person and number in pronouns: A feature-geometric analysis. *Language*, 78, 482–526. https://doi.org/10.1353/lan.2002.0158

Lardiere, D. (1998a). Case and tense in a 'fossilized' steady state. *Second Language Research*, 14(1), 1–26. https://doi.org/10.1191/026765898674105303

Lardiere, D. (1998b). Disassociating syntax from morphology in a divergent L2 end-state grammar. *Second Language Research*, 14(4), 359–375. https://doi.org/10.1191/026765898672500216

Lardiere, D. (2009). Some thoughts on the contrastive analysis of features in second language acquisition. *Second Language Research*, 25(2), 173–227. https://doi.org/10.1177/0267658308100283

Lardiere, D. (2017). Detectability in feature reassembly. In S. M. Gass, P. Spinner, & J. Behney (Eds.), *Salience in Second Language Acquisition* (pp. 41–63). New York, NY: Routledge.

Li, M. C. (2012). The acquisition of tense and agreement by early child second language learners. Unpublished PhD dissertation, University of Illinois at Urbana-Champaign.

Rothman, J., & Slabakova, R. (2018). The state of the science in generative SLA and its place in modern second language studies. *Studies in Second Language Acquisition*, 40(2), 417–442. https://doi.org/10.1017/S0272263117000134

Slabakova, R. (2008). *Meaning in the second language*. New York/Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110211511

Slabakova, R. (2016). *Second language acquisition*. Oxford: Oxford University Press.

White, L. (1989). *Universal grammar and second language acquisition*. Amsterdam: John Benjamins. https://doi.org/10.1075/lald.1

White, L. (2003). *Second language acquisition and universal grammar*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511815065

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

Yang, C. (2018). A formalist perspective on language acquisition. *Linguistic Approaches to Bilingualism*, 8(6), 665–706.

Address for correspondence

Roumyana Slabakova
University of Southampton
Faculty of Humanities
Building 65, Avenue Campus
Highfield, Southampton, SO17 1BF
United Kingdom

R.Slabakova@soton.ac.uk
https://orcid.org/0000-0002-5839-460X

Publication history

Date received: 5 July 2018
Date accepted: 15 August 2018

Learning rules versus learning items

Peter Svenonius
University of Tromsø – The Arctic University of Norway

From my perspective as a theoretical syntactician, Yang's contributions to the study of language acquisition (as summarized in Yang, 2018) appear to be game-changing. The Tolerance Principle seems to have substantial support and broad application; the fact that it and the Sufficiency Principle make clear predictions which are easily tested suggests that we will quite soon learn more about how broad those applications are and what kinds of refinements are needed.

Something which I would like to see better integrated into this kind of work is a theory of the hypothesis space entertained by children's Linguistic Acquisition Devices (LADs) in their pursuit of grammar. Yang describes children positing rules, which is convenient for his exposition, but most theoretical syntacticians, including myself, do not think of grammar in terms of language-specific rules. Rather, there are language-specific lexical items which interact with universal principles and operations. Constraints on the 'format' of lexical items therefore define the hypothesis space available to the learner (see Rizzi, 2017). We can go further and separate two aspects of language-specific lexical items, namely the 'head' which is an atom of syntactic computation and its morphophonological 'exponent.'

In these terms, there isn't a rule of English dictating that *-(e)d* is added to the verb stem to get a past tense. Instead, there are two lists of items to be learned. One includes the suffixal exponent *-(e)d* and its competitors for expressing T[past], the special exponents in *saw*, *made*, *fought*, and so on.

The other list which must be learned is the list of inflectional heads of the clause, characterized as T for tense, including the general properties common to all T heads and the specific properties of the head T[past] of which *-(e)d* is an exponent. T[past] combines with the verbal head V as a suffix in affirmative declaratives, but not in negative clauses or in interrogatives. Other T elements, such as modals and auxiliary verbs, share many properties with T[past] but not others, for example modals are like T[past] in inverting with the subject in interrogatives, but modals are not affixal.

https://doi.org/10.1075/lab.18067.sve
Linguistic Approaches to Bilingualism 8:6 (2018), **782–786**.
ISSN 1879-9264 | E-ISSN 1879-9272

In these terms, the learning of the 'rule' that *-(e)d* can be suffixed to form a past tense is better cast as the acquisition of an ***item*** T[past] which is a bound (suffixal) head, among other properties, and another ***item*** *-(e)d* which is a default exponent of T[past], where 'default' means simply that it is not specified for what verb stems it combines with. That more fully specified competitors 'block' the default when their conditions are met is a general principle which need not be learned (Kiparsky, 2005).

The effect is the same as what is described by Yang: *-(e)d* can be recognized as the default (i.e., underspecified) when it achieves a certain numerical dominance. My replacement of a simply stated rule by a combination of lists and principles might seem like splitting of terminological hairs, but in other cases I believe it matters, as I illustrate here first with Norwegian Object Shift, and then with determiner-noun combinatorics in English.

In Norwegian, the verb precedes the object, just as in English, and follows 'middle field' adverbs (those which are neither at the right nor the left periphery); the word order in a typical embedded clause is English-like, as illustrated in (1a) (unlike English, there is no "*do*-support" in case of negation).

(1)	at	de	ikke	spiste	den
	that	they	not	ate	it

In main clauses the finite verb appears in second position. This has the effect that the verb can be separated from the object by adverbs and by the subject, as in (2).

(2)	a.	Da	spiste	måsen	ikke	pølsa.
		then	ate	the.seagull	not	the.hotdog
	b.	Da	spiste	måsen	ikke	noe
		then	ate	the.seagull	not	anything
	c.	Da	spiste	måsen	ikke	DEN
		then	ate	the.seagull	not	that

The object shift rule is simple to state:

(3) If the object is a noncontrastive unstressed definite pronoun (like 'it' or 'me' or 'her') in a clause where the main verb is in second position, then the pronoun precedes middle-field adverbs; otherwise it follows them.

Object shift is illustrated in (4). It is generally considered obligatory in Northern and Eastern Norwegian (see Bentzen, 2014 on the dialectal distribution); failure of object shift, as in (5), is typically rejected in elicitation and is infrequent in speech (Bentzen, 2014; Bentzen, Anderssen, & Waldman, 2013). Some sense of focus or contrast is regularly implied by lack of shift, as indicated by the gloss 'that' in (2c), and is usually accompanied by stress, as indicated by the use of capitals in (2c).

(4) Da spiste måsen den ikke.
then ate the.seagull it not

(5) **Da spiste måsen ikke den.
then ate the.seagull not it

However, children overproduce sentences like (5) even as late as the age of seven. Anderssen, Bentzen, and Rodina, (2012) report an experiment in which 4–5-year-olds failed to shift 81% of the time, and 6–7-year-olds failed to shift 30.6% of the time, in contexts considered obligatory by adults.

Something must prevent young children from formulating (3). This would be explained if the hypothesis space available to the LAD is constrained in terms of features on heads, as envisioned by Rizzi. The LAD cannot entertain (3) because the grammar is not stated in terms of stress or linear order of elements. Instead, the variable placement of arguments illustrated in (1–2) and (4) indicates formal properties of functional heads, in this case a functional head in the T domain which attracts definite pronouns under certain conditions related to verb placement.

The discussion in terms of rules also bears on Yang's (2018) discussion of the combinatorial freedom of determiners and nouns. Because of the history of the argument going back to Valian (1986), Yang writes as if children learn a rule like (6).

(6) Singular count N can combine with either *a* or *the*

Yang notes that this rule has no chance of surviving the numerical demands of the Sufficiency Principle if the full inventory of nouns in use is considered. He notes, however, that, in the Adam corpus, 83 of the mother's ***100 most frequent nouns*** are observed to satisfy (6). For $N = 100$, N/log(N) is 21.71, so just 78 positive examples would allow (6) to be adopted as a rule, and 83 is more than sufficient, if only the most frequent nouns are considered.

But most syntacticians don't think that languages have rules like (6). Instead, languages like English (and Norwegian, and Italian, and German) require determiners in noun phrases, or certain noun phrases, and children learn this by the age of three or earlier (e.g., Guasti, Gavarró, de Lange, & Caprin, 2008). Alongside *a* and *the*, possessive pronouns like *my* and *his* and indefinite quantifiers like *some*, *any*, and *many* are also determiners, in English.

The vast majority of English nouns in contexts of use appear with determiners of one kind or another. The exceptions fall into well-defined classes; names normally occur with no determiner. Mass nouns like *water* and plurals like *children* can also appear without one, in the sorts of indefinite contexts that require *a* for singular count nouns. Most syntacticians therefore posit a null counterpart to *a* for mass and plural nouns, allowing the generalization that noun phrases headed by common nouns in English require determiners. If there is also a null definite determiner for names (Longobardi, 1994), then all noun phrases require determiners, when they function as arguments (mainly subjects, objects, complements of prepositions, and possessors).

So the real generalization has to do with the construction of argument noun phrases, which crucially involve a category D (for determiner), at least in languages like English. It is not about a possibility available to nouns, but an obligatory property of noun phrases. This can insightfully be stated in the properties of the heads comprising the noun phrase.

The LAD needs to determine, not how many kinds of noun appear with both *a* and *the*, but rather, which elements have the 'D' property necessary for the establishment of an argument? There are no lexically irregular forms of indefiniteness or definiteness in English, so learning the exponents is trivial. But learning the determiner heads behind the exponents is a lot like learning the T heads underlying tenses, modals and auxiliaries. Elements like *a*, *the*, *many*, *some*, *a lot*, and *another* could in principle be the same category or different categories. The LAD needs to identify and categorize these elements – some as determiners of category D, some as adjectives of category A, and some as quantifiers of some other category. Some of these elements might be systematically coupled with a silent determiner, and the LAD will have to determine when that is the best analysis. In that case we are counting not how many of hundreds of nouns appear with *a* and *the*, but how many of a smaller number of adnominal elements share which properties with each other.

Clearly, none of this can be taken to support the claim that determiner-noun combinations are learned as unanalyzed collocations. Yang's statistical analysis shows that children's output is exactly what would be expected if they are acquiring the English system accurately. (It should be noted, too, that data from bilingual children dramatically illustrates the failure of a rote learning account, since bilingual children quite commonly combine determiners from one language with nouns from the other (e.g., Cantone, 2007, p. 66). Examples like *un schiff* 'a (Italian) ship (German)' and *il fernsehn* 'the (Italian) TV (German)' are typical, from ages 2;1 and 2;4, respectively, from Cantone, (2007, p. 152).)

It is to their credit that Yang's proposals transcend differences among frameworks and approaches. However, I think that independent shortcomings of a rule-based approach to grammatical competence carry over into applications of Yang's theory which assume them, and that a more constrained approach to variation will give better results.

Acknowledgements

Thanks to Kristine Bentzen for discussion of the object shift facts and their implications.

References

Anderssen, M., K. Bentzen, & Y. Rodina. (2012). Topicality and complexity in the acquisition of Norwegian object shift. *Language Acquisition* 19(1), 39–72. https://doi.org/10.1080/10489223.2012.633844

Bentzen, K. (2014). Object Shift. *Nordic Atlas of Linguistic Structures (NALS)*. Tekstlab, Oslo.

Bentzen, K., M. Anderssen, & C. Waldman. (2013). Object shift in spoken Mainland Scandinavian. *Nordic Journal of Linguistics* 36(2), 115–151. https://doi.org/10.1017/S0332586513000218

Cantone, K. F. (2007). *Code-switching in bilingual children*. Springer, Dordrecht.

Guasti, M. T., A. Gavarró, J. de Lange, & C. Caprin. (2008). Article omission across child languages. *Language Acquisition* 15, 89–119. https://doi.org/10.1080/10489220801937346

Kiparsky, P. (2005). Blocking and periphrasis in inflectional paradigms. *Yearbook of Morphology* 2005, 113–135. https://doi.org/10.1007/1-4020-2900-4_5

Longobardi, G. (1994). Reference and proper names: A theory of N-movement in syntax and logical form. *Linguistic Inquiry* 25(4), 609–665.

Rizzi, L. (2018). On the format and locus of parameters: The role of morphosyntactic features. *The Linguistic Review* 41(3–4), 159–190.

Valian, V. (1986). Syntactic categories in the speech of young children. *Developmental Psychology* 22(4), 562–579. https://doi.org/10.1037/0012-1649.22.4.562

Yang, C. (2018). A formalist perspective on language acquisition. *Linguistic Approaches to Bilingualism*, 8(6), 665–706.

Address for correspondence

Peter Svenonius
Department of Language and Culture
Faculty of Humanities, Social Sciences and Education
UiT – The Arctic University of Norway
N-9037 Tromsø
Norway

peter.svenonius@uit.no
https://orcid.org/0000-0001-5043-8543

Publication history

Date received: 10 July 2018
Date accepted: 23 August 2018

Formalist modeling and psychological reality

Eva Wittenberg and Ray Jackendoff
University of California, San Diego / Tufts University

Linguists use rules to explain patterns in language. They are our version of mathematicians' equations – we try to make them elegant and simple. They compress the mess. They are sometimes not entirely accurate; but when they're not, we can call the deviations exceptions, and perhaps we can state a separate rule to adjust for them.

To explain patterns in people's *behavior*, however, many linguists in recent years have transitioned from elegant and simple rules to elegant and simple mathematical models: Surprisal, to predict and explain sentence processing results (e.g., Levy, 2008); Information Density, to understand language production (e.g., Jaeger, 2010); and Bayes' Rule, to formalize pragmatic reasoning processes (e.g., Goodman & Frank, 2016). It is hard to overestimate how influential these approaches have become. They too are sometimes not entirely accurate; but when they're not, we can call it unexplained variation, and perhaps we can tweak the model's parameters, weights, or priors to adjust for it.

Charles Yang (2018) attempts to build a bridge between rules and equations in language acquisition: explaining both the patterns of learners' grammars and their behavior in comprehension and production. He takes linguistic analysis and mathematical modeling equally seriously, as one should, and he does not shy away from hard problems with complex data.

We are enthusiastic about these goals. Our commentary concerns a general question which could be asked about *any* of the mathematical approaches above, as well as Yang's: How do the constructs of these models correspond to what is going on *in the human mind*? Are these models compatible with what we know (or at least think) human language learners do? We recognize that any theory requires a certain degree of idealization, but we wonder whether the idealizations in Yang's models are empirically realistic. Here are a number of issues where these concerns arise.

https://doi.org/10.1075/lab.18077.wit
Linguistic Approaches to Bilingualism 8:6 (2018), **787–791**.
ISSN 1879-9264 | E-ISSN 1879-9272

1. Where do candidate grammars and rules originate?

In Yang's variational learning model, the learner is testing candidate grammars against linguistic input. If a grammar successfully parses the input, it is rewarded: its probability is incremented. If it fails to parse the input, it is penalized: its probability is decremented.[1] In the Tolerance Principle model, words in the input are checked to see whether they conform to candidate rules. In both cases, the question is: What is the source of these candidates?

One standard response has been to claim that the space of candidate grammars is determined by an innate Universal Grammar. Yet, given the grammatical uniqueness of practically every language ever studied, the number of grammars of possible human languages is far too large to build into a brain, in such a way that an acquisition process like the variation learning model could prune all these candidates down to a single choice.

One way of triaging the possibilities is to propose that a grammar is determined by setting the values of a smallish set of innate parameters (Chomsky, 1981; Baker, 2001) one of which, the Null Subject Parameter, is discussed by (Yang, 2018). Learning a grammar then amounts to setting the values of the parameters. Parameters have however fallen on hard times. As Newmeyer (2004) observes, there prove to be too many picky little syntactic differences among languages and too many thoroughly idiosyncratic constructions, for instance, Culicover's (1999) "syntactic nuts" to be able to maintain the prospect of defining possible languages in terms of a small set of innate parameters.

The only alternative we can envision is that candidate rules and grammars are themselves constructed by the learner. Of course, this too requires an innate basis that defines the space of possible rules of grammar. However, if the innate basis is generative, this space, even if highly constrained, could still be unlimited – as in fact appears to be the case. On this story, an important part of the acquisition process is then how the learner uses the innate basis to construct appropriate candidate rules in response to the input. The choices between candidates may still be governed by Yang's variational learning, but first the candidates have to be built. Perhaps here is where ideas from usage-based learning may be useful.

1. However, what counts as penalizing or rewarding "a grammar" as a whole is not clear. In practice, in illustrating the model, Yang in effect deals with competition among individual *rules* – a computationally far more tractable task.

2. Do children really use variational learning?

An old chestnut from the acquisition literature (McNeill, 1966, p. 69) leads us to question the premises of variational learning.

(1) a. Child: Nobody don't likes me.
b. Mother: Say "Nobody likes me."
c. Child: Nobody don't likes me.

Presumably the inputs to the child have all been of the form (1b), which in variational learning would reward a candidate grammar that coincides with the adults', while penalizing the double negative grammar. Yet the child's utterances betray the (temporary) victory of the double negative grammar, in the absence of compatible input. This seems to violate the premises of the variational learning model: how could an incorrect grammar ever achieve high probability, such that children persist in their mistakes in the face of counterevidence?

3. How is compliance tested?

In order to determine whether an input is in compliance with a grammar (in variational learning) or with a rule (in Tolerance Principle learning), the learner has to mentally represent the input in a form that can be tested against the grammar or rule. What is this form? And since it has to be established prior to a choice of rules, what principles assign it?

4. What happens when a rule is chosen? What happens when it's dethroned?

Yang's mechanism presumes that if a base can undergo a rule, the output is not listed in the lexicon (e.g., *walk* is stored but *walked* is not). Yet, the output of at least some instances of the rule had to be listed at one point in time, in order to lead to the construction of the rule in the first place (see issue 1 above). Moreover, if the candidate rule does not work out because one has discovered too many exceptions, its putative outputs are now themselves exceptions and have to be listed. In other words, the rule's outputs must be listed long enough to motivate the rules, only to be thrown away, and then resurrected if the rule happens to be a failure.

5. How does the mathematics map into a mechanism?

More globally, we find implausible many aspects of Yang's proposed mechanism: first, that in lexical access the brain performs a serial search over the lexicon (Yang, 2016); second, that it computes (the moral equivalent of) a numerical count of the lexical items that potentially undergo each rule; third, that it constructs and regularly updates a ranked list of exceptions to each rule; and fourth, that it also counts the number of bases for which there is currently no information as to whether they undergo the rule or not. To be clear: Yang does not specifically discuss when or how these operations take place, but he apparently envisions his approach as based in mental reality – something the brain must do. But how?

As we mentioned at the beginning, these are open questions not only for Yang's models, but for any mathematical model put forth in psycholinguistics. How are these mathematical constructs operationalized? For instance, whenever there are global calculations over a corpus, such as determining the probability of a grammar or a Bayesian conditional probability, or subtracting the number of exceptions from the number of instances, or calculating surprisal based on the frequency of *n*-grams, there is an implicit assumption that these factors have cognitive and/or neural correlates. Frequency in a corpus is no doubt intended as a proxy for something like strength of neural activation, but it remains to be seen whether the mathematics as a whole can be recast in these terms.

An easy way out of this conundrum is to disclaim responsibility for positing a mechanism, on the grounds that one is simply proposing a "computational" theory in Marr's (1982) sense, a theory which delineates the abstract structure of the cognitive domain. How the structure is stored and processed falls under an "algorithmic" or "implementational" theory, taken to be an entirely different enterprise. While we understand the virtues of computational theories, we reject the notion that they can be insulated from theories of the mechanism, just as we reject the parallel notion that linguistic competence can be insulated from performance – especially in the study of language acquisition.

References

Baker, M. C. (2001). *The atoms of language*. New York: Basic Books.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

Culicover, P. W. (1999). *Syntactic nuts: Hard cases, syntactic theory, and language acquisition*. Oxford: Oxford University Press.

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, 20(11), 818–829. https://doi.org/10.1016/j.tics.2016.08.005

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, 61(1), 23–62. https://doi.org/10.1016/j.cogpsych.2010.02.002

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Marr, D. (1982). *Vision*. San Francisco: Freeman.

McNeill, D. (1966). *Developmental psycholinguistics*. In: *The genesis of language*, F. Smith and G. Miller (Eds.), 15–84. Cambridge: MIT Press.

Newmeyer, F. J. (2004). Against a parameter-setting approach to typological variation. *Linguistic Variation Yearbook*, 4(1), 181–234. https://doi.org/10.1075/livy.4.06new

Yang, C. (2016). *The price of linguistic productivity: How children learn to break rules of language*. MIT Press, Cambridge, MA.

Yang, C. (2018). A formalist perspective on language acquisition. *Linguistic Approaches to Bilingualism*, 8(6), 665–706.

Address for correspondence

Eva Wittenberg
University of California
Department of Linguistics
9500 Gilman Drive
La Jolla, CA 92093-0108
United States of America

ewittenberg@ucsd.edu
https://orcid.org/0000-0002-3188-6145

Co-author details

Ray Jackendoff
Tufts University
Center for Cognitive Studies
115 Miner Hall
Medford, MA 02155
United States of America

ray.jackendoff@tufts.edu
https://orcid.org/0000-0002-1358-9200

Publication history

Date received: 14 August 2018
Date accepted: 20 August 2018

Input effects on the development of I-language in L2 acquisition

Noriaki Yusa
Miyagi Gakuin Women's University

In his keynote article, Yang (2018) presents us with a welcome amalgam of domain-specific Universal Grammar (UG) and experience-based inductive learning in first (L1) and second language (L2) acquisition, arguing that the interaction of UG and domain-general learning mechanisms contributes to the development of internalized linguistic knowledge (I-language). Yang's variational learning model assumes that language acquisition proceeds gradually by setting parameters on the basis of the probabilistic promotion or demotion of each grammatical option. The Tolerance Principle asserts that a productive rule can be tolerated so long as the number of its exceptions does not exceed its critical threshold. Yang's article succeeds in incorporating input effects into the framework of generative approaches to L1 and L2 acquisition, which is an impressive feat.

My comments are concerned with the role of input in the development of I-language in L2 acquisition. First, the variational learning mechanism predicts that the developmental time course of parameter-resetting in L2 acquisition correlates with the frequency of the unambiguous evidence in the input that L2 learners are exposed to. Despite its high input frequency, however, English subject-verb agreement as in "John *walks* to school every day" is a well-known phenomenon that often poses enormous challenges for even advanced users of English. For example, Lardiere (2007) reports the famous case study of Patty, an L1 Chinese user of English living in the United States for more than two decades, who consistently omits agreement on thematic verbs. Judging from her mastery of other grammatical phenomena including negation, pronominal nominative case and the relative ordering of adverbs and verbs, she seems to have acquired the knowledge of tense related to those phenomena but she cannot access the morphological spell-outs of this knowledge (see also Domínguez & González Alonso; Slabakova; this volume). Yang (2018) states that "the L2 learners' apparent inability to use a simple rule fully productively is because they know *too many words*" (p. 694). It remains unclear how larger vocabulary raising the threshold for productive use of a simple rule will

https://doi.org/10.1075/lab.18081.yus
Linguistic Approaches to Bilingualism 8:6 (2018), **792–796**.
ISSN 1879-9264 | E-ISSN 1879-9272

help to explain the difficulty for L2 learners to use the simple rule of subject-verb agreement consistently. It is also unclear how Yang's model based on spontaneous production data can capture the knowledge of agreement not reflected in L2 learners' production data.

Second, in addition to the quantity of input, the quality of input matters in the ultimate attainment of native-like processing of L2 acquisition (Piske & Young-Scholten, 2009). There is growing body of evidence that social interaction is a critical factor influencing adult L2 acquisition (Verga & Kotz, 2013, Yusa et al., 2017) For example, fMRI research by Yusa et al. (2017) shows that learning a foreign language (syntax of Japanese sign language) through a deaf signer in a classroom resulted in a different pattern of activation of the left inferior frontal gyrus (IFG) than learning it through the identical amount of input via a DVD recording of the classroom situation: learning accompanied by changes in brain functions is not triggered solely by linguistic input such as DVDs, but is enhanced by social interaction. If the activation in the left IFG is involved in syntactic processing in L1 and L2 (Perani & Abutalebi, 2005), Yusa et al.'s result implies that learning L2 syntax in a richer social context may well lead to closer native-like attainment of L2 processing. It also suggests that despite methodological constraints, future qualitative research on the type and quality of input in L2 acquisition will reveal a more realistic picture of how the $e < N/ln\ N$ rule in the Tolerance Principle may apply, depending on the type of input.

The last comment concerns the effect of input frequency on the development of I-language in L2 learners. Yang discusses the L2 acquisition of null subjects by referring to the published literature showing that the acquisition of null subjects is similar between L1 and L2 acquisition in the sense that null subjects in Chinese and Italian appear earlier than overt expletive subjects in English. For example, "L1 English learners of Japanese have close-to-native command of null subjects…" while "even near-native L2 learners of English fail to consistently use the expletive subject" (2018, p. 689). Even if it is the case that L1 English learners of Japanese can correctly use null subjects in Japanese, this does not mean that they have the same I-language as Japanese native speakers regarding the deeper properties of Japanese null subjects. Reviewed from this perspective, Yang's research on null subjects is based on the traditional assumption that null subjects in Japanese are unpronounced pronominals (*pro*). However, recent syntactic research reveals that null subjects in radical *pro*-drop East Asian languages such as Japanese and Korean are not *pro*'s but are derived from argument ellipsis (Oku, 1998; Saito, 2007), mainly based on the empirical fact that Japanese null subjects allow a sloppy reading.

(1) a. Taro-wa [[jibun-no teian]-ga saiyoosareru]-to
Taro-TOP self-GEN proposal-NOM will.be.accepted-COMP
omot-te-iru.
think-PROG-PRES
'Taro$_i$ thinks his$_i$ proposal will be accepted.'
b. Hanako-mo [[*e*] saiyoosareru]-to omot-te-iru.
Hanako-also will.be.accepted-COMP think-PROG-PRES
Lit. 'Hanako also thinks [*e*] will be accepted.' (strict/sloppy)

(2) Hanako-mo sore-ga saiyoosareru-to omot-te-iru.
Hanako-also it-NOM will.be.accepted-COMP think-PROG-PRES
'Hanako also thinks it will be accepted.' (strict/*sloppy)

The embedded null subject in (1b) can be interpreted either as Taro's proposal (strict reading) or as Hanako's proposal (sloppy reading). The availability of the latter construal is unexpected if the position is occupied by a null pronoun, *pro*, since pronouns do not allow a sloppy reading, as exemplified in (2). If null arguments in radical *pro*-drop East Asian languages are not *pro*'s but are derived from argument ellipsis, it does not follow from the close-to-native performance of Japanese null subjects by L1 English speakers reported in Yang's paper that they use the same argument ellipsis strategy in the subject position as native speakers of Japanese do. It might be the case that they have only learned the visible or distributional properties such as the availability of null subjects in Japanese but that they have not acquired non-visible deep properties such as the availability of a sloppy reading with null subjects in Japanese. In fact, recent L2 studies by Yamada and Miyamoto (2017) and M. Yusa (2018) show, contrary to Yang's observation, that L1 English learners of Japanese have difficulty acquiring the knowledge of a sloppy reading in the subject position. Note here that the availability of a sloppy reading in the Japanese null subject is an empirical fact, whether the null argument in Japanese is derived from argument ellipsis or not. Therefore, to show that L2 learners have acquired the knowledge of null subjects in Japanese, it is imperative to show that they have also acquired the knowledge of a sloppy reading in the position. The inductive distributional analysis of L2 input, which keeps track of the number of null subjects in Japanese input, may reasonably induce L2 Japanese learners to use null subjects frequently in Japanese. The frequent use of Japanese null subjects does not, however, mean that input frequency of null subjects has affected the development of I-language regarding a sloppy reading in L2 Japanese learners.

To conclude, Yang discusses the role of input (second factor) in significant depth in L1 and L2 acquisition by combining UG (first factor) and the Tolerance Principle (third factor) (Chomsky, 2005). It may be customary to advocate either a nativist approach or a usage-based approach to language acquisition, but the combination of the two approaches as exemplified in Yang's paper seems to be more fruitful to get a real picture of language acquisition by using the abundant amount of empirical data available from both approaches. My comments are not intended to refute Yang's approach, but rather to clarify his approach by raising three questions about input effects on I-language in L2 acquisition.

Funding

This work was supported by a grant from the Japan Society for the Promotion of Science JP16K13266, JP17H02364 awarded to Noriaki Yusa.

References

Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry*, 36(1), 1–22. https://doi.org/10.1162/0024389052993655

Lardiere, D. (2007). *Ultimate attainment in second language acquisition: A case study*. Mahwah, NJ: Lawrence Erlbaum Associates.

Oku, S. (1998). A theory of selection and reconstruction in the minimalist perspective. PhD dissertation, University of Connecticut.

Perani, D., & Abutalebi, J. (2005). The neural basis of first and second language processing. *Current Opinion in Neurobiology*, 15(2), 202–206. https://doi.org/10.1016/j.conb.2005.03.007

Piske, T., & Young-Scholten, M. (Eds.). (2009). *Input matters in SLA*. Bristol: Multilingual Matters.

Saito, M. (2007). Notes on East Asian argument ellipsis. *Language Research*, 43, 203–227.

Verga, L., & Kotz, S. A. (2013). How relevant is social interaction in second language learning? *Frontiers in Human Neuroscience*, 7, 550. https://doi.org/10.3389/fnhum.2013.00550

Yamada, K., & Miyamoto, Y. (2017). On interpretation of null arguments in L2 Japanese by European non-pro-drop and pro-drop language speakers, *Journal of the European Second Language Association*, 1(1), 73–89. https://doi.org/10.22599/jesla.18

Yang, C. (2018). A formalist perspective on language acquisition. *Linguistic Approaches to Bilingualism*, 8(6), 665–706.

Yusa, M. (2018). Acquisition of Japanese null arguments by second language learners. MA thesis. The Ohio State University.

Yusa, N., Kim, J., Koizumi, M., Sugiura, M., & Kawashima, R. (2017). Social interaction affects neural outcomes of sign language learning as a foreign language in adults. *Frontiers in Human Neuroscience*, 11, 115. https://doi.org/10.3389/fnhum.2017.00115

Address for correspondence

Noriaki Yusa
Miyagi Gakuin Women's University
Department of English
9-1-1, Sakuragaoka
Aoba-ku, Sendai, 981-8557
Japan

yusa@mgu.ac.jp
https://orcid.org/0000-0001-6818-7180

Publication history

Date received: 20 August 2018
Date accepted: 21 August 2018

REPLY

Some consequences of the Tolerance Principle

Charles Yang

The commitment to a formalist theory of language acquisition seems to have resonated with the commentators. In what follows, I will discuss and expand on some of the central issues surrounding the Tolerance Principle (TP).

1. Clarifications and corrections

The book-length treatment of the TP (Yang, 2016) would have provided a fuller background and assuaged some of the commentators' concerns. For instance, both **Wittenberg and Jackendoff** and **Kapatsinski** are incredulous that language could operate like a serial search model, the algorithmic foundation of the TP – because, they argue, the brain is parallel. But is a parallel brain really incompatible with serial behavioral effects? My (presumably parallel) brain can memorize and recite the digits of π in a strictly linear sequence. And there are numerous serial effects in the study of cognition that are the product of a parallel brain: the scanning memory model of Sternberg (1969), the linear search model of number representation and processing (Gallistel et al., 1992; Brannon et al., 2001), not to mention Weber's Law (Gibbon, 1977). More to the point, a serial model is simply a better account of lexical processing. A picture (actually two) is worth a thousand words.

The results are based on lexical decision data from almost 40,000 words (Balota et al., 2007). Factors affecting reaction time (word length, orthographic neighborhood size, etc.) have been controlled ("residualized"; see Lignos, 2013 for details). Word rank (bottom) clearly provides a better fit than word frequency (top). Incidentally, even frequency-based accounts always use the logarithm of frequency, which brings it surreptitiously close to rank.

I do agree with Wittenberg and Jackendoff that one should always pursue multi-level explanations (in the sense of Marr). I have done so myself (Yang, 2016, 76–78, Yang, 2017) while criticizing Bayesian rational analysis models that explicitly disavow psychological reality – which, confusingly, Kapatsinski

https://doi.org/10.1075/lab.00022.yan
Linguistic Approaches to Bilingualism 8:6 (2018), **797–809**.
ISSN 1879–9264 | E-ISSN 1879–9272

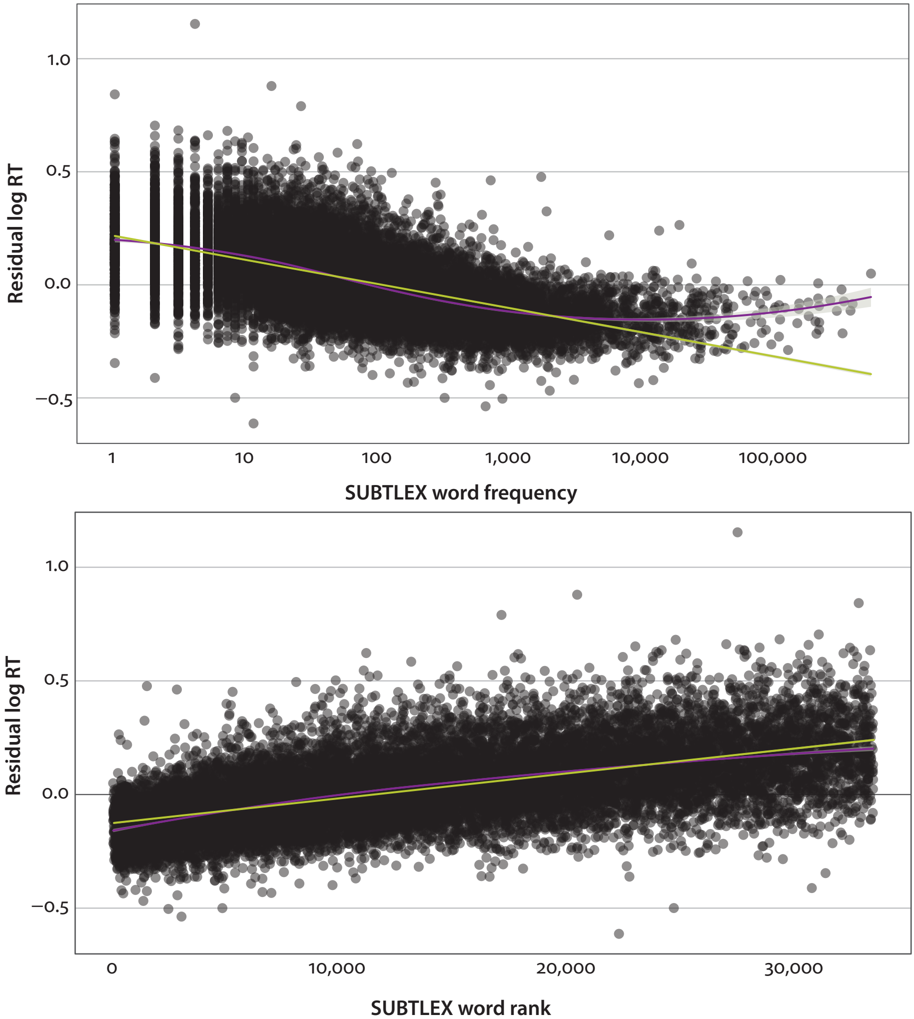

Figure 1. Comparison of word frequency and rank as predictor of lexical decision time. From Lignos (2013)

embraces as an article of faith despite his concern for psychological grounding. The appropriate response should be to develop a neural theory of serial effects, rather than disregarding serial effects from behavioral studies just because 'the brain is parallel', especially when these commentators don't even offer insight on how the parallel brain implements *parallel* effects.

Kapatsinski seems to have read the book but only selectively. He completely ignored my thorough cross-linguistic review of morphological acquisition, which shows unambiguously that productivity is a categorical effect (Xu & Pinker, 1995).

Rather, he prefers his own study of adult artificial language learning where the subjects, on average, produced a gradient score on a Wug test and concludes that productivity must be gradient. But we already know that adults do not approach (artificial) language learning tasks in the categorical way that children do (Hudson Kam & Newport, 2005); more in Section 2. And it's important to recall that the original Wug study (Berko, 1958) already demonstrated the task-specific differences between children and adults. The subjects were presented with novel verbs such as spow and bing, which are similar to existing irregular verbs (blow-blew, sing-sang). Only one of the 86 children tested produced bang (no said said spew, and only spowed was produced). Adults, however, are far more willing to produce irregularized forms, which is likely to a task effect, since there have been virtually no such cases in the natural history of English verbs (Anderwald, 2013; Yang, 2016).

Goldberg also read the book but doesn't seem to have understood it. She wonders how one learns the prefix *pre-* as in *pre-Watergate, pre-Trump,* etc. According to her, since children will learn many nouns and proper names but presumably relatively few will appear with *pre-*, the productivity threshold would not be reached and the prefix cannot be learned. This is a perverse reading of the TP and the general problem of inference. Just learning a word does not force the learner to evaluate all conceivable ways in which the word is be used. The past tense *-ed* can only be learned when (enough) verbs have appeared in past tense; the suffix wouldn't even be entertained when the child hears and uses a verb in the present. In the acquisition study of the dative constructions summarized in the target article, ditransitive verbs such as *slip* are not included because every instance of it in the child-directed input corpus is an intransitive form (e.g., *They slipped).* By the same token, not every noun or proper name – thank goodness – would reach the notoriety of *Watergate* and *Trump,* so *truck, milk,* and *Abby* do not factor into the calibration of *pre-* at all. Goldberg seems to believe that if a form (e.g., "pre-tortilla") *can* be used, it *must* be used. This clearly doesn't follow but it does explain her persistent appeal to indirect negative evidence (e.g., Boyd & Goldberg, 2011), the contrapositive of the above: If a form is *not* used, it must *not* be grammatical – which fails empirically as well (Yang, 2015). In the rest of her commentary, Goldberg summarizes her own proposal: "(P)reviously witnessed partially-abstracted exemplars cluster together in our hyper-dimensional representational space for language, forming a massively interrelated dynamic system (a construct-i-con), which is simply an expanded version of what has long been recognized to be needed for our knowledge of words (the lexicon)" (p. 729). I have no idea what this means. While I'm quite prepared for someone to show the TP to be wrong – so long as as they know how to use it right – it should really be replaced with a better equation, not vague analogies and metaphoric allusions.

I am pleased that two prominent usage-based researchers, **Gries** and **Rowland,** agree with my call for methodological rigor: Gries has made similar pleas and Rowland even gives me a Popperian endorsement. The TP, so far as I can tell, is an example of the learning mechanism that has been viewed as the central goal of usage-based researchers: "a single mechanism responsible both for generalization, and for restricting these generalizations to items with particular semantic, pragmatic, phonological (and no doubt other) properties (Ambridge & Lieven, 2011, p. 267)." A useful common ground.

But their defense of the usage-based position is unconvincing. **Gries** proposed a log odds ratio measure which shows the frequency of *give me* is higher than "expected by chance", and is thus "at least compatible with the notion that *gimme* might be a unit" (p. 735). But neither point is correct. It is true that *me* follows *give* more frequently than "by chance", with "chance" understood as "other words" (Gries; Table 1). But the only way to establish productivity is statistical independence; i.e., the frequency of *give me* can be predicted from the frequency of *give* and that of *me*. The "other words" do not matter. Furthermore, what if give me is indeed abnormally frequent? It still doesn't follow that give me is a holistic unit. Frequency and compositionality are in principle independent of each other. Gries seems to uphold the idea that whole-unit frequency effects – if real, though not in the present case – automatically counts against compositionality. This is a dogma from the past tense debate as pointed out in the target article; see Yang (2002), Taft (2004), Fruchter and Marantz (2015), Regel et al. (2015) for acquisition, processing, and neurological evidence for compositional approaches to whole-unit frequency effects.

Rowland does not directly challenge the statistical findings of my determiner work (Yang, 2013a) but brings up the study of Pine et al. (2013). In some subsamples of child language, children are assessed to have a lower overlap score than adults. Rowland faults me for not discussing this result; here is why. The Pine et al. (2013) method is biased, and generally undervalues the productivity of the smaller sample, which is usually the child corpus because adults talk more. One can develop an analytical diagnosis of the bias – too complex to summarize here (see Yang & Valian, 2018) – but the problem can also be revealed by a minimum sanity check, on samples produced by (adult) speakers whose knowledge of the determiners is not in question. Doing so would have shown Mary MacWhinney to be (absurdly) less productive than Brian (the curator of CHILDES) in the MacWhinney corpus: Brian just talked a lot more. Once again, quantitative results are interpreted at face value, and methods that produce (preferred) results are deployed without validation.

2. Children, adults, and vocabularies

Is there any continuity between L1 and L2, not to mention those cases – "atypical" development, simultaneous and sequential early multiple acquisition, heritage language and attrition, etc. – that lie in between? This is obviously too large of question and my goal is much more specific. I propose that the TP is operative for adult language learning, and its smaller-is-better property, that rules are easier than learn when the relevant vocabulary is smaller, can account for adult's evident deficiency in comparison to children.

How do we show that the TP is used by adults at all? There is *prime facie* evidence that suggests otherwise. For example, in a series of studies (summarized in Schuler, 2017), subjects learn artificial languages where rules have various levels of exceptions. Children follow the TP nearly categorically but adults generally match the token frequencies of the available forms in the stimuli.

It remains unclear why adults probability match while children prefer categorical rules, a difference found in other domains of learning and decision making (e.g., Weir, 1964; Derks & Paclisanu, 1967). Yet there are at least two reasons to believe that the TP holds for learners of all ages. Theoretically, the central assumption of the Tolerance Principle, namely the Elsewhere Condition, is not known to degrade across development. Empirically, there is evidence that adults, and late child learners, can learn rules extremely well. First, a significant portion of English derivational morphology is acquired quite late (Tyler & Nagy, 1989; Jarmulowicz, 2002), presumably because it involves advanced vocabulary that comes with literacy and education. Second, L2 learners *can* form productive rules in a manner similar to L1 learners (White & Genesee, 1996). Yang and Montrul (2017) provide an extensive review of L2 acquisition of the English dative constructions. These constructions are informative because their grammatical properties are obscure and most English teachers, I'd imagine, would never offer lessons on the distribution of *donate*. But L2 learners also go through the stages of over-generalization and retreat like L1 learners, and they gradually refine the phonological and semantic restrictions on the constructions over time, with advanced learners showing native-like grammaticality judgment (Jäschke & Plag, 2016)

The commentators are correct to stress the complexity of L2 acquisition. **Paradis** highlights the individual differences in L2 that cannot be attributed to "language-level" factors such as word frequencies. **Dominíguez and González Alonso, Montrul,** and **Yusa** point out that the input for L2 acquisition is filtered through the learner's L1. The target article recognizes these complications. For instance, I chose adult Italian learners of English to demonstrate the presence of a topic-drop grammar (á la Chinese and Japanese; see also Yusa, this volume), which is neither in the speaker's L1 (pro-drop) or L2 (obligatory subject), thereby providing unambiguous evidence for Full Access. Similarly, the analysis of determiner productivity in L2 shows comparable statistical results for Italian and Punjabi speakers despite the differences in their L1 determiner systems, which should address **Dimroth**'s concerns. And the TP, with its focus on vocabulary composition, is well equipped to handle both language- and individual-level differences. The relative ease of French past tense acquisition noted by Paradis would follow my account of why the English plural suffix *-s* is learned earlier than the past tense *-ed* (fewer exceptions; Yang, 2016, 4.1.3). And the onset of rule productivity for English-learning children Adam, Eve, and Abe can be predicted by their vocabulary acquisition (Yang, 2016, 4.1.2).

Under the TP, the effect of L1 on L2 is formally no different from (purely) L1 acquisition. It is trivially true that the child doesn't learn everything they hear; otherwise they would learn 50,000 words by the age of two. But just saying the input is filtered (**Biberauer, Perkins & Lidz**) does not solve anything; one still needs to understand how rules come out of the "intake". I think there is little prospect in a general theory of filtering because the input may be reduced by a virtually unlimited range of factors: cognitive limitations in children, L1 influence in L2 adults, a kid overly obsessed with Lego, an ESL student who Facebooks rather than paying attention in class. But the learner's vocabulary composition, both L1 and L2, can be estimated and the TP will makes clear claims about grammatical rules no matter how filtering works.

On the matter of vocabulary, several commentators (**De Cat, Dimroth, Slabakova**) question my take on less-is-more. I should have been more clear: while a smaller vocabulary does make rule learning easier, it still needs to be large enough for the rule to be learnable (e.g., enough regular verbs to overcome the irregulars). Thus a younger learner may not be better than an older one at everything: I have already discussed the gradual refinement of the English dative constructions because the requisite vocabulary can only be built up over time, so older learners are "better" than younger ones. Thus, **Dimroth**'s interesting study that child L2 learners perform better than L1 learners on German verbal morphology is perfectly compatible with the TP: although a fine grained corpus analysis is necessary, the complexity of the German system would seem to require a substantial vocabulary which an older child may acquire faster. And it is definitely *not* the case that bilingual children would learn rules faster than monolingual learners, contrary to **De Cat**'s understanding: reduced input as in the case of bilingual acquisition will reduce the vocabulary necessary to support rule productivity, which is exactly what Marchman et al. (2010) find.

A related question, raised by **De Cat, Dimroth,** and **Kapatsinski** in somewhat different ways, concerns the completion of rule learning. Since the TP requires a great majority of words to follow a rule to ensure productivity, waiting too long before coming to a decision (i.e., with a large N) would render every rule unproductive because of the data sparsity (Yang, 2013b). The answer comes in two ways, both suggesting that the learner will stop looking and "freeze" the rules in place at a value of N no more than a few hundred. Empirically, a three-year-old's vocabulary size is no more than just around 1,000 (Hart & Risley, 1995). This is at an age where the core grammar (word order, inflectional morphology, etc. though not everything) is already solidly in place. Thus, productivity decisions can, and must, be made when N is quite modest. It is important to stress that the learning limit of N is not a function of age: the full details of the dative constructions are learned quite late but the value of N for the verbs is probably no more than 100; see also the discussion of L2 datives above. Conceptually, as I discussed elsewhere (Yang, 2016, 76ff, Yang, 2018), the TP has been surprisingly, and unreasonably, effective, especially because the numerical assumptions in its derivation are almost never strictly true (and no one even bothers checking). It seems that children somehow keep track of two quantities and compute their relations. It is difficult to envision high-precision calculation for large N's although the neural implementation of something like the TP is completely unknown.

3. Learnability and the theory of grammar

The last set of comments comes from theoretical linguists or acquisition researchers who make a strong ontological commitment to theoretical linguistics. Some worry whether the TP has gone too far in the other direction, without paying sufficient attention to the representational and other constraints in the grammar (**De Cat, Dominíguez & González Alonso, Perkins & Lidz, Roeper, Slabakova**).

My general approach is to have as little UG as possible (Berwick & Chomsky, 2016). The application of the TP has been, by design, based on what can be described as plausible generalizations about the data without making (unnecessary) theoretical commitments about how such generalizations are to be stated. For instance, "add *-ed* to verbs to form past tense" can be stated either "in the lexicon" or "in the syntax" – a matter of fierce theoretical controversy but the bean counting of N, e, and θ_N is all the same. The TP provides a lower bound on what is distributionally learnable from data. If this approach is successful, then explanatory adequacy no longer resides in the intricacies of theory-internal apparatus or principles and constraints specific to language (Yang et al., 2017).

This will invite skepticism. **Perkins and Lidz** believe my theory fails to take developmental constraints into account. They also question children's ability to detect semantic properties, e.g., caused-possession in the dative constructions. For them, it is the syntax that helps the learner to narrow down the semantics (syntactic bootstrapping; Gleitman, 1990). But they don't seem to realize that the TP already provides a *developmental* theory of syntactic bootstrapping: syntax does help with semantics but syntax has to be learned.[1] Table 3 of the target article shows how the vocabulary of dative verbs, and thus the double-object construction, grows over time. Let's grant that children can't "observe" the meaning of verbs such as *promise* and *guarantee* without syntax (although Perkins & Lidz only offer assertions to this effect without evidence). The syntax for bootstrapping can be formed when the vocabulary is small and contains only "easy" words (Gleitman et al., 2005) – *give, feed, hand, show, bring* – whose meanings *are* observationally learnable (Trueswell et al., 2016). This provides the TP-sanctioned inductive basis that caused possession is encoded in the double-object structure, with which children can decode the meanings of *promise* and *guarantee*.

Perkins and Lidz are also concerned that my approach may "miss important generalizations about language acquisition" (p. 743) such as "(I)f a language has two ditransitive constructions, the one expressing caused possession is always the one in which the goal c-commands the theme … And, children seem to know this link despite a severe poverty of evidence" (p. 746). I fail to see the relevance. How does knowing the goal c-commanding the theme help learn that *donate* cannot appear in the double-object construction but *assign* can? Never mind the supposed generalization is false: Middle English (Visser, 1963) and modern Scandinavian languages to varying degrees, allow both the goal-theme and theme-goal word order. The same holds for the suggestion that rule learning may be aided by features and other formal structures (**Biberauer, Dominínez & González Alonso, Slabakova).** Perhaps productive and unproductive processes are indeed differentiated representationally but that is clearly the result of learning not a theory of learning, e.g., which words belong where on the hierarchy, which features become general and "abstract" and which are conservative and lexically specific. In this vein, **Svenonius** raised the problem of object shift in Norwegian, where children fail to consistently shift in obligatory contexts. The distribution of shifted objects can be described in different theoretical frameworks with some more surface-oriented that others but the learning problem is the same and has already been subjected to a TP analysis (Anderssen et al., 2012, 57). The number of shifted object pronominals is only a small subset of all (10/39): not shifting is "productive" and children must memorize those that do. Failing to shift consistently is expected because exceptions may be regularized as in the familiar case of English past tense. Similarly, the obligatory use of determiners in languages like Italian needn't follow from the property of some null head – and one would need a story

1. I thank Lila Gleitman for discussions of this matter over some funky blue bread pudding.

of why it is *not* available for English – but can be learned distributionally from input (Ceolin, 2018).

But Svenonius's general message is important: what are the "constraints on the 'format' of lexical items therefore define the hypothesis space" (p. 781). UG won't provide a complete set of the primitives to structure the hypothesis space. It is inconceivable that the noun classes in Bantu, the classifier system in Japanese, and the past tense rules for the irregular verbs in English are all carved out of the innate universal template. More likely, these linguistic categories are established because children can discover, using something like the TP, formal correspondences that relate them. A case in point is the "telecommunication" subclass of dative verbs, which is surely not an innate semantic class but one established on their participation in a formal structure namely the double object construction. The child is probably innately primed to organize the categories in a combinatorial system ("features"), which may follow from Merge and perhaps other general principle of system organization (e.g., the particulate principle; Studdert-Kennedy, 1998).

4. Conclusion

The TP provides a new division of labor between what can be learned and what needs to be built in. As **Rothman and Chomsky** point out, this can eliminate "arbitrary stipulations of parameter values" (p. 765) and provides an account of the idiosyncractic properties of particular grammars without overburdening the biological requirement for language. Indeed, the minimalist approach (Berwick & Chomsky, 2016) encourages a return to an earlier, abductive, framework of language acquisition: "Having selected a permissible hypothesis, he can use inductive evidence for corrective action, confirming or disconfirming his choice. Once the hypothesis is sufficiently well confirmed, the child knows the language defined by this hypothesis; consequently, his knowledge extends enormously beyond his experience" (Chomsky, 1968, p. 80). The TP determines whether a hypothesis is "sufficiently well confirmed".

It seems appropriate to end with the concluding remarks from Rothman and Chomsky, who quote Chomsky (1995): "The field is changing rapidly under the impact of new empirical materials and theoretical ideas. What looks reasonable today is likely to take a different form tomorrow. … Whether these steps are on the right track or not, of course, only time will tell" (p. 9). This will take a collective endeavor from many theoretical and empirical angles as the commentators have helpfully made clear.

References

Ambridge, B. & Lieven, E. V. (2011). *Child language acquisition: Contrasting theoretical approaches*. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511975073

Anderssen, M., Bentzen, K., & Rodina, Y. (2012). Topicality and complexity in the acquisition of norwegian object shift. *Language Acquisition*, 9(1), 39–72. https://doi.org/10.1080/10489223.2012.633844

Anderwald, L. (2013). Natural language change or prescriptive influence?: Throve, dove, pled, drug and snuck in 19th-century American English. *English World-Wide*, 34(2), 146–176.

Armstrong, S. L., Gleitman, L. R., & Gleitman, H. (1983). What some concepts might not be. *Cognition*, 13(3), 263–308. https://doi.org/10.1016/0010-0277(83)90012-4

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, 39(3), 445–459. https://doi.org/10.3758/BF03193014

Berko, J. (1958). The child's learning of English morphology. *Word*, 14(2–3), 150–177. https://doi.org/10.1080/00437956.1958.11659661

Berwick, R. C. & Chomsky, N. (2016). *Why only us: Language and evolution*. Cambridge, MA: MIT Press.

Biberauer, T. (2018). Less is More: On the Tolerance Principle as a manifestation of Maximize Minimal Means. *Linguistic Approaches to Bilingualism*, 8(6), 707–711.

Boyd, J. K. & Goldberg, A. E. (2011). Learning what not to say: The role of statistical preemption and categorization in a-adjective production. *Language*, 87(1), 55–83. https://doi.org/10.1353/lan.2011.0012

Brannon, E. M., Wusthoff, C. J., Gallistel, C., & Gibbon, J. (2001). Numerical subtraction in the pigeon: Evidence for a linear subjective number scale. *Psychological Science*, 12(3), 238–243. https://doi.org/10.1111/1467-9280.00342

Ceolin, A. (2018). Explaining cross-linguistic differences in article omission through an acquisition model. In Bertolini, A. B. & Kaplan, M. G., (Eds.), *Proceedings of the 42nd annual Boston University Conference on Language Development*, pp. 100–113, Somerville, MA: Cascadilla Press.

Chomsky, N. (1968). *Language and mind*. Harcourt, Brace and World.

Chomsky, N. (1995). *The minimalist program*. Boston: MIT Press.

Derks, P. L. & Paclisanu, M. I. (1967). Simple strategies in binary prediction by children and adults. *Journal of Experimental Psychology*, 73(2), 278. https://doi.org/10.1037/h0024137

De Cat, C. (2018). Evaluating Yang's algorithms: An outline. *Linguistic Approaches to Bilingualism*, 8(6), 712–716.

Dimroth, C. (2018). Input and the acquisition of productive grammatical knowledge: Vocabulary size as missing link? *Linguistic Approaches to Bilingualism*, 8(6), 717–721.

Domínguez, L. & González Alonso, J. (2018). What is the role of L1 representations in a grammar-input model of L2 acquisition? *Linguistic Approaches to Bilingualism*, 8(6), 722–726.

Fruchter, J. & Marantz, A. (2015). Decomposition, lookup, and recombination: MEG evidence for the Full Decomposition model of complex visual word recognition. *Brain and Language*, 143, 81–96. https://doi.org/10.1016/j.bandl.2015.03.001

Gallistel, C. R. & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, 44(1–2), 43–74. https://doi.org/10.1016/0010-0277(92)90050-R

Gibbon, J. (1977). Scalar expectancy theory and weber's law in animal timing. *Psychological Review*, 84(3), 279. https://doi.org/10.1037/0033-295X.84.3.279

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, 1(1), 3–55. https://doi.org/10.1207/s15327817la0101_2

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). Hard words. *Language Learning and Development*, 1(1), 23–64. https://doi.org/10.1207/s15473341lld0101_4

Goldberg, A. (2018). The sufficiency principle hyperinflates the price of productivity. *Linguistic Approaches to Bilingualism*, 8(6), 727–732.

Gries, S. Th. (2018). Mechanistic formal approaches to language acquisition: Yes, but at the right level(s) of resolution. *Linguistic Approaches to Bilingualism*, 8(6), 733–737.

Hart, B. & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Paul H Brookes Publishing, Baltimore, MD.

Hudson Kam, C. L. & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, 1(2), 151–195. https://doi.org/10.1080/15475441.2005.9684215

Jarmulowicz, L. (2002). English derivational suffix frequency and children's stress judgements. *Brain and Language*, 81(1–3), 192–204. https://doi.org/10.1006/brln.2001.2517

Jäschke, T. & Plag, I. (2016). The dative alternation in German-English interlanguage. *Studies in Second Language Acquisition*, 38, 485–521. https://doi.org/10.1017/S0272263115000261

Kapatsinski, V. (2018). On the intolerance of the Tolerance Principle. *Linguistic Approaches to Bilingualism*, 8(6), 738–742.

Lidz, J. & Perkins, L. (2018). The importance of input representations. *Linguistic Approaches to Bilingualism*, 8(6), 743–748.

Lignos, C. (2013). Modeling words in the mind. PhD thesis, University of Pennsylvania.

Marchman, V. A., Fernald, A., & Hurtado, N. (2010). How vocabulary size in two languages relates to efficiency in spoken word recognition by young spanish-english bilinguals. *Journal of Child Language*, 37(4), 817–840. https://doi.org/10.1017/S0305000909990055

Montrul, S. (2018). Learning a Second Language Takes More than Math. *Linguistic Approaches to Bilingualism*, 8(6), 749–752.

Paradis, J. (2018). Language-level input factors are not enough to explain child bilingual acquisition. *Linguistic Approaches to Bilingualism*, 8(6), 753–757.

Pine, J. M., Freudenthal, D., Krajewski, G., & Gobet, F. (2013). Do young children have adultlike syntactic categories? Zipf's law and the case of the determiner. *Cognition*, 127(3), 345–360. https://doi.org/10.1016/j.cognition.2013.02.006

Regel, S., Opitz, A., Müller, G., & Friederici, A. D. (2015). The past tense debate revisited: Electrophysiological evidence for subregularities of irregular verb inflection. *Journal of Cognitive Neuroscience*, 27(9), 1870–1885. https://doi.org/10.1162/jocn_a_00818

Roeper, T. (2018). Grammar acquisition and grammar choice in the variationist model. *Linguistic Approaches to Bilingualism*, 8(6), 758–763.

Rothman, J. & Chomsky, N. (2018). Towards eliminating arbitrary stipulations related to parameters: Linguistic innateness and the variational model. *Linguistic Approaches to Bilingualism*, 8(6), 764–769.

Rowland, C. (2018). The principles of scientific inquiry. *Linguistic Approaches to Bilingualism*, 8(6), 770–775.

Schuler, K. (2017). The acquisition of productive rules in child and adult language learners. PhD thesis, Georgetown University, Washington, D.C.

Schütze, C. T. (2005). Thinking about what we are asking speakers to do. In Kepser, S. & Reis, M., (eds.), *Linguistic evidence: Empirical, theoretical, and computational perspectives*, pages 457–485. Mouton de Gruyter, Berlin. https://doi.org/10.1515/9783110197549.457

Slabakova, R. (2018). Back to our roots. *Linguistic Approaches to Bilingualism*, 8(6), 776–780.

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

Studdert-Kennedy, M. (1998). The particulate origins of language generativity: from syllable to gesture. In Hurford, J., Studdert-Kennedy, M., & Knight, C., (eds.), *Approaches to the evolution of language*, pages 202–221. Cambridge University Press, Cambridge.

Svenonius, P. (2018). Learning rules versus learning items. *Linguistic Approaches to Bilingualism*, 8(6), 781–786.

Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *The Quarterly Journal of Experimental Psychology*, 57A(4), 745–765. https://doi.org/10.1080/02724980343000477

Trueswell, J. C., Lin, Y., Armstrong, B., Cartmill, E. A., Goldin-Meadow, S., & Gleitman, L. R. (2016). Perceiving referential intent: Dynamics of reference in natural parent-child interactions. *Cognition*, 148, 117–135. https://doi.org/10.1016/j.cognition.2015.11.002

Tyler, A. & Nagy, W. (1989). The acquisition of English derivational morphology. *Journal of Memory and Language*, 28(6), 649–667. https://doi.org/10.1016/0749-596X(89)90002-8

Visser, F. T. (1963). *An historical syntax of the English language*. Brill Archive.

Weir, M. W. (1964). Developmental changes in problem-solving strategies. *Psychological Review*, 71(6), 473. https://doi.org/10.1037/h0041785

White, L. & Genesee, F. (1996). How native is near-native? the issue of ultimate attainment in adult second language acquisition. *Second Language Research*, 12(3), 233–265. https://doi.org/10.1177/026765839601200301

Wittenberg, E. & Jackendoff, R. (2018). Formalist modeling and psychological reality. *Linguistic Approaches to Bilingualism*, 8(6), 787–791.

Xu, F. & Pinker, S. (1995). Weird past tense forms. *Journal of Child Language*, 22(3), 531–556. https://doi.org/10.1017/S0305000900009946

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

Yang, C. (2013a). Ontogeny and phylogeny of language. *Proceedings of the National Academy of Sciences*, 110(16), 6324–6327. https://doi.org/10.1073/pnas.1216803110

Yang, C. (2013b). Who's afraid of George Kingsley Zipf? Or: Do children and chimps have language? *Significance*, 10(6), 29–34. https://doi.org/10.1111/j.1740-9713.2013.00708.x

Yang, C. (2015). Negative knowledge from positive evidence. *Language*, 91(4), 938–953. https://doi.org/10.1353/lan.2015.0054

Yang, C. (2016). *The price of linguistic productivity: How children learn to break rules of language*. MIT Press, Cambridge, MA.

Yang, C. (2017). Rage against the machine: Evaluation metrics in the 21st century. *Language Acquisition*, 24(2), 100–125. https://doi.org/10.1080/10489223.2016.1274318

Yang, C. (2018). *A user's guide to the Tolerance Principle. Manuscript*. University of Pennsylvania (ling.auf.net/lingbuzz/004146).

Yang, C., Crain, S., Berwick, R. C., Chomsky, N., & Bolhuis, J. J. (2017). The growth of language: Universal grammar, experience, and principles of computation. *Neuroscience and Biobehavioral Reviews*, 81(Part B), 103–119. https://doi.org/10.1016/j.neubiorev.2016.12.023

Yang, C. & Montrul, S. (2017). Learning datives: The tolerance principle in monolingual and bilingual acquisition. *Second Language Research*, 33(1), 119–144. https://doi.org/10.1177/0267658316673686
Yang, C. & Valian, V. (2018). Determiners and grammars. Submitted.
Yusa, N. (2018). Input effects on the development of I-language in L2 acquisition. *Linguistic Approaches to Bilingualism*, 8(6), 792–796.

Author's address

Charles Yang
Department of Linguistics and Computer Science
University of Pennsylvania
3401 Walnut Street 315C
Philadelphia, PA 19104, 215-898-7849

charles.yang@ling.upenn.edu