Some Consequences of the Tolerance Principle

Charles Yang

charles.yang@ling.upenn.edu

September 30, 2018

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 & Jackdoff** 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.
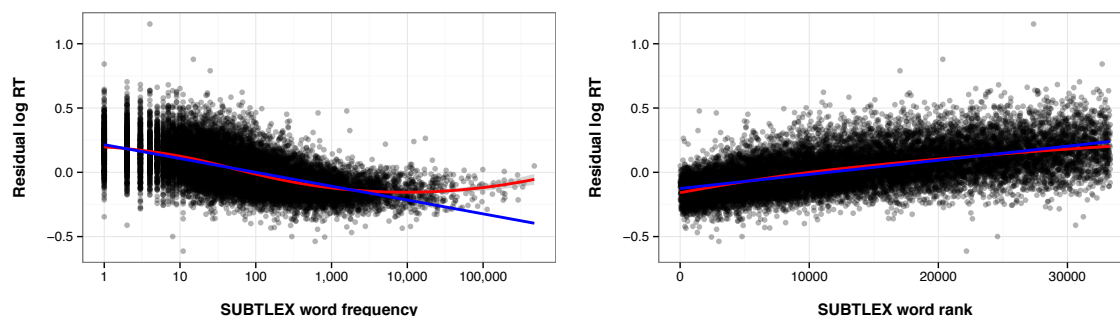

Figure 1: Comparison of word frequency and rank as predictor of lexical decision time. From Lignos (2013).

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 (right panel) clearly provides a better fit than word frequency (left). 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 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 head 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 (Berko, 1958, Xu and Pinker, 1995). Rather, he prefers his own study of adult artificial language learning where the subjects provided gradient scores on a rating task 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 and Newport, 2005); more in Section 2. Worse, Kapatsinski's rating task *guarantees* a gradient conclusion because the task itself is gradient. This should have been obvious since Armstrong et al. (1983) famously demonstrated that 7 is rated as a better odd number than 23. Adults may well give a non-zero rating for *boze* as the past tense of *bize* and *fludge* for *flidge*. When multiple choices are made available, subjects will use them (Parducci and Perrett, 1971, Schütze, 2005), but does anyone actually say *boze* and *fludge*?

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 and 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 inter-related 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)." 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 and Lieven, 2011, 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 from chance", and is thus "at least compatible with the notion that *gimme* might be a unit." 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 and 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 and 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 and 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 and 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 and 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. **Domíníguez & 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), 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 and 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**, **Domínguez & Alonso**, **Perkins & Lidz**, **Roeper**, **Slabakova**).

My general approach is to have as little UG as possible (Berwick and 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 & 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 & Lidz are also concerned that my approach may "miss important generalizations about language acquisition" 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." 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**, **Domínínez & 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 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 that define the hypothesis space." 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

---

[1] I thank Lila Gleitman for discussions of this matter over some funky blue bread pudding.

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 & Chomsky** point out, this can eliminate "arbitrary stipulations of parameter values" and provides an account of the idiosyncractic properties of particular grammars without overburdening the biological requirement for language. Indeed, the minimalist approach (Berwick and 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, p80). The TP determines whether a hypothesis is "sufficiently well confirmed".

It seems appropriate to end with the concluding remarks from Rothman & Chomsky: "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." This will take a collective endeavor from many theoretical and empirical angles as the commentators have helpfully made clear.

References

Ambridge, B. and Lieven, E. V. (2011). *Child language acquisition: Contrasting theoretical approaches*. Cambridge University Press, Cambridge.

Anderssen, M., Bentzen, K., and Rodina, Y. (2012). Topicality and complexity in the acquisition of norwegian object shift. *Language Acquisition*, 19(1):39–72.

Armstrong, S. L., Gleitman, L. R., and Gleitman, H. (1983). What some concepts might not be. *Cognition*, 13(3):263–308.

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., and Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, 39(3):445–459.

Berko, J. (1958). The child's learning of English morphology. *Word*, 14(2–3):150–177.

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

Boyd, J. K. and 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.

Brannon, E. M., Wusthoff, C. J., Gallistel, C., and Gibbon, J. (2001). Numerical subtraction in the pigeon: Evidence for a linear subjective number scale. *Psychological Science*, 12(3):238–243.

Ceolin, A. (2018). Explaining cross-linguistic differences in article omission through an acquisition model. In Bertolini, A. B. and Kaplan, M. G., editors, *Proceedings of the 42nd annual Boston University Conference on Language Development*, pages 100–113, Somerville, MA. Cascadilla Press.

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

Derks, P. L. and Paclisanu, M. I. (1967). Simple strategies in binary prediction by children and adults. *Journal of Experimental Psychology*, 73(2):278.

Fruchter, J. and 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.

Gallistel, C. R., Gelman, R., et al. (1992). Preverbal and verbal counting and computation. *Cognition*, 44(1-2):43–74.

Gibbon, J. (1977). Scalar expectancy theory and weber's law in animal timing. *Psychological review*, 84(3):279.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, 1(1):3–55.

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., and Trueswell, J. C. (2005). Hard words. *Language Learning and Development*, 1(1):23–64.

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

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

Hudson Kam, C. L. and 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.

Jarmulowicz, L. (2002). English derivational suffix frequency and children's stress judgements. *Brain and Language*, 81(1–3):192–204.

Jäschke, T. and Plag, I. (2016). The dative alternation in German–English interlanguage. *Studies in Second Language Acquisition*, 38:485–521.

Lignos, C. (2013). *Modeling words in the mind*. PhD thesis, University of Pennsylvania.

Marchman, V. A., Fernald, A., and 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.

Parducci, A. and Perrett, L. F. (1971). Category rating scales: Effects of relative frequency of stimulus values. *Journal of Experimental Psychology*, 89(2):427–452.

Pine, J. M., Freudenthal, D., Krajewski, G., and 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.

Regel, S., Opitz, A., Müller, G., and 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.

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. and Reis, M., editors, *Linguistic evidence: Empirical, theoretical, and computational perspectives*, pages 457–485. Mouton de Gruyter, Berlin.

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., and Knight, C., editors, *Approaches to the Evolution of Language*, pages 202–221. Cambridge University Press, Cambridge.

Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *The Quarterly Journal of Experimental Psychology*, 57A(4):745–765.

Trueswell, J. C., Lin, Y., Armstrong, B., Cartmill, E. A., Goldin-Meadow, S., and Gleitman, L. R. (2016). Perceiving referential intent: Dynamics of reference in natural parent–child interactions. *Cognition*, 148:117–135.

Tyler, A. and Nagy, W. (1989). The acquisition of English derivational morphology. *Journal of Memory and Language*, 28(6):649–667.

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.

White, L. and 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.

Xu, F. and Pinker, S. (1995). Weird past tense forms. *Journal of Child Language*, 22(3):531–556.

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

Yang, C. (2013b). Who's afraid of George Kingsley Zipf? Or: Do children and chimps have language? *Significance*, 10(6):29–34.

Yang, C. (2015). Negative knowledge from positive evidence. *Language*, 91(4):938–953.

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.

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., and Bolhuis, J. J. (2017). The growth of language: Universal grammar, experience, and principles of computation. *Neuroscience and Biobehavioral Reviews*, 81(Part B):103 – 119.

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

Yang, C. and Valian, V. (2018). Determiners and grammars. Submitted.