Negative knowledge from positive evidence

Charles Yang

Language, Volume 91, Number 4, December 2015, pp. 938-953 (Article)

Published by Linguistic Society of America

NEGATIVE KNOWLEDGE FROM POSITIVE EVIDENCE

Charles Yang

University of Pennsylvania

Why can't we say *the asleep cat*? There is a class of adjectives in English, all of which start with a schwa (e.g. *afraid*, *alone*, *asleep*, *away*, etc.), that cannot be used attributively in a prenominal position. A frequently invoked strategy for the acquisition of such negative constraints in language is to use indirect negative evidence. For instance, if the learner consistently observes paraphrases such as *the cat that is asleep*, then the conspicuous absence of *the asleep cat* may be a clue for its ungrammaticality (Boyd & Goldberg 2011). This article provides formal and quantitative evidence from child-directed English data to show that such learning strategies are untenable. However, the child can rely on positive data to establish the distributional similarities between this apparently idiosyncratic class of adjectives and locative particles (e.g. *here*, *over*, *out*, etc.) and prepositional phrases. With the use of an independently motivated principle of generalization (Yang 2005), the ungrammaticality of attributive usage can be effectively extended to the adjectives in question.*

Keywords: language acquisition, indirect negative evidence, distributional learning, corpus linguistics, exceptions, computational linguistics, constructions

1. Introduction. A theory of language and language acquisition should provide a broad account of the speaker's linguistic competence. It must explain the expressions that the speaker can produce as well as the absence of expressions that the speaker has not and in fact cannot produce because they are prohibited by universal or language-specific constraints.

The negative aspect of language, or what NOT to say, has long been recognized as a central problem in the study of language acquisition. The most prominent discussion centers on what is known as baker's paradox (Baker 1979): How do children delineate the possible from the impossible based only on positive linguistic examples? Consider the two dative constructions in 1a.

(1) a. John told Bill the story.
John told the story to Bill.
b. John donated the painting to the museum.
*John donated the museum the painting.

How do children know that the double object construction is ungrammatical for verbs like *donate* (1b), even though they encounter plenty of positive instances of interchangeability between the two dative constructions, as in 1a? The absence of negative evidence to children (Brown & Hanlon 1970, Braine 1971, Bowerman 1988, Marcus 1993) means that they cannot rely on direct feedback. At the same time, they cannot assume unattested linguistic forms to be ungrammatical in general, for that would rule out the productive and infinite use of language.

The problem of learning negative constraints has led to a sizable body of literature and advanced our understanding of language acquisition (see Baker 1979, Mazurkewich & White 1984, Berwick 1985, Fodor & Crain 1987, MacWhinney 1987, Fodor 1989, Pinker 1989, Randall 1990, Pesetsky 1995, Inagaki 1997, Campbell & Tomasello 2001, Conwell & Demuth 2007, etc.). In this article, I wish to show that, at least in some cases, the exceptional patterns of language use are only apparent. To acquire them still raises interesting questions for language acquisition, but it does not lead the learner,

* I would like to thank Ben Bruening, Kyle Gorman, Julie Anne Legate, two anonymous referees, and the editors of *Language* for their helpful comments and suggestions.

or the theorist, into any logical or empirical quagmire. Negative knowledge of language can be acquired on the basis of positive evidence alone, as is illustrated with the case of the so-called *a*-adjectives.

It has long been observed (Bolinger 1971, Bouldin 1990, Beard 1995, Huddleston & Pullum 2001, Larson & Marušič 2004, Cinque 2010) that there is a class of English adjectives that can be used predicatively but not attributively in a prenominal position. These adjectives start with an unstressed schwa (*a*) and have thus acquired the label *A*-ADJECTIVES.

(2) a. The cat is asleep. ??the asleep cat
b. The boss is away. ??the away boss
c. The dog is awake. ??the awake dog
d. The child is alone. ??the alone child
e. The troops are around. ??the around troops

How do children learn that *a*-adjectives cannot go in attributive position? In a recent article, Boyd and Goldberg (2011) claim that these properties of *a*-adjectives are idiosyncratic and require the strategy of STATISTICAL PREEMPTION in language acquisition. According to these authors, the ungrammaticality of attributive usage (e.g. *the asleep cat*) is prevented by paraphrases such as *the sleeping cat* or a relative clause *the cat that is asleep*, similar to the blocking effect in morphology (*held* over *holded*) and syntax (Poser 1992).

In this article, an alternative learning strategy that makes use of positive evidence to learn negative exceptions is considered. Following and extending analyses put forward by Salkoff (1983), Larson and Marušič (2004), Coppock (2008), Bruening (2011a,b), and others, I first present morphological and syntactic evidence that *a*-adjectives pattern with locative particles, such as *up*, *out*, *over*, *here*, and *there*, and prepositional phrases. Such distributional equivalence ultimately enables the learner to extend the prohibition on attributive usage from locative particles and prepositional phrases to *a*-adjectives. In §3, I discuss formal and empirical problems with the statistical preemption approach, while offering an alternative interpretation of Boyd and Goldberg's (2011) experimental results. Quantitative analyses of a 4.3-million-word corpus of child-directed English are provided in §4. I show that the learner has sufficient evidence to establish the distributional properties of *a*-adjectives under an independently motivated learning principle for generalization (Yang 2005, 2016). The article concludes with a discussion of the necessity of and challenge with using corpus data in the psychological study of language.

2. *A*-ADJECTIVES ARE NOT ATYPICAL. Historically, many of the *a*-adjectives were derived from prepositional phrases (Long 1969), which do not appear before the noun in a noun phrase. While the etymology of words is unlikely to be available to the child learner, there is still synchronic evidence (Rauh 1993, Stvan 1998, Bruening 2011a,b) that reveals these adjectives' PP-like characteristics.

First, locative particles *present*, *out*, *over*, *on*/*off*, *up*, *here*/*there*, and so on are words that, like *a*-adjectives, also resist attributive use in a prenominal position.

(3) a. The chairperson is present. ??the present chairperson (spatial sense)
b. The receptionist is out. ??the out receptionist (spatial sense)
c. The batter is up. ??the up batter
d. The matches are over. ??the over matches
e. The delivery is here. ??the here delivery

In this regard, both *a*-adjectives and locative particles pattern like prepositional phrases (PPs).

(4) a. The ball is out of sight. ??the out-of-sight ball
b. The dog is behind the fence. ??the behind-the-fence dog
c. The singers are at ease. ??the at-ease singers
d. The marbles are in the jar. ??the in-the-jar marbles

In addition, the attributive use of *a*-adjectives improves when they are further modified (as in 5; Huddleston & Pullum 2001), and the same holds for locative particles (6) and prepositional phrases (7).

(5) a. ?the nocturnally awake cat
b. ?a frequently away parent
c. ?an aware, amused look (Huddleston & Pullum 2001:559)

(6) a. ?the almost here train
b. ?the never present advisor
c. ?an up-and-down experience

(7) a. ?the always on-time teacher
b. ?an angry but on-fire hitter
c. ?a still-in-service shuttle

As noted by Salkoff (1983:299) and Coppock (2008:181), *a*-adjectives share a well-defined morphological structure; they are not an arbitrary list of adjectives that happen to share an initial schwa. Indeed, the ungrammaticality of attributive usage appears associated not with *a*-adjectives per se but with the aspectual prefix *a*-, as shown in the novel adjectives in 8.

(8) a. The tree is abud with green shoots.
??An abud tree is a beautiful thing to see.
b. The water is afizz with bubbles.
??The afizz water was everywhere.

Larson and Marušič (2004) observe that all *a*-adjectives can be decomposed into the prefix *a*- and a stem that is typically free but sometimes bound (e.g. *aghast* with *ghast* appearing in *ghastly*, *afraid* with *fraid* in *fraidy*, *aware* with *ware* in *beware*). The list in 9 is taken from their article (p. 270) with a few of my own additions; none appears to be acceptable in attributive use.

(9) abeam, ablaze, abloom, abuzz, across, adrift, afire, aflame, afraid, agape, aghast, agleam, aglitter, aglow, aground, ahead, ajar, akin, alight, alike, alive, alone, amiss, amok, amuck, apart, around, ashamed, ashore, askew, aslant, asleep, astern, astir, atilt, awake, aware, awhirl, away, awash

By contrast, the attributive restriction disappears if the adjective consists of the schwa *a*- (i.e. nonprefix) and a nonstem (10) or a pseudo-stem (11) that does not contribute to the meaning of the composite.

(10) a. the above examples
b. the aloof professor
c. the alert student
d. the astute investor

(11) the acute problem

Of course, the language acquisition problem does not go away under the morphological approach to *a*-adjectives. First, learners need to recognize that the *a*-stem combination forms a well-defined set of adjectives, as in 9, that is structurally distinct from the phonologically similar but morphologically simplex adjectives, as in 10–11. Second, and more importantly, they need to learn that the *a*-adjectives thus formed cannot be used attributively, which is the main problem at hand.

I do not review all of the similarities between *a*-adjectives and locative particles discussed by previous researchers. Some of the diagnostics, such as those based on conjunction and *-ly* suffixation to establish the syntactic categories of *a*-adjectives, have proven less than conclusive (Bruening 2011b, Goldberg 2011). More critically, these diagnostics use UNGRAMMATICAL examples: while invaluable to the theorist in uncovering the complexity of linguistic knowledge, they do not have an obvious role to play in the course of language acquisition. For example, the improvement of attributive use upon further modification (5–7) is not at all attested in our child-directed English corpus and is therefore not helpful for the learner in establishing the properties of *a*-adjectives. Likewise, attested usage examples from the Web, which previous researchers have used to study the properties of *a*-adjectives, cannot be assumed to be available to the learner without an evaluation of the child-directed language data. We return to the issue of corpus data and language acquisition in the concluding section of this article.

One of the diagnostics, however, has proven very robust and, as we see in §4 below, is abundantly attested in the input to English-learning children. A class of adverbs such as *right*, *well*, *far*, and *straight*, which express the meaning of intensity or immediacy, can be used to modify *a*-adjectives. Following Bruening (2011b), I collectively refer to these structures as *right*-type modification.

(12) a. I was well/wide awake at 4 AM.
b. The race leader is well ahead.
c. The baby fell right/sound asleep.
d. You can go right ahead.
e. The guards are well aware (of the danger).

To be sure, probably not all *a*-adjectives may be used with *right*-type modification: *I became well/right afraid* is ungrammatical for most English speakers who were surveyed, although one can find attested examples in very large corpora. But such adverbial modification cannot appear at all with typical adjectives (13), while it is compatible with both locative particles (14) and PPs (15).

(13) a. *The car is right/straight/well new/nice/red.
b. *The politician is right/straight/well annoying/amazing/available.
(14) a. The referee was right here/there.
b. The cat came straight out.
c. The rocket soared right up.
d. The answer was wide off.
e. The arrow was shot well over.
(15) a. The referee was right in the penalty box.
b. The cat ran straight out of the house.
c. The rocket soared right across the sky.
d. The answer was wide off the mark.
e. The arrow was shot well over the fence.

Despite these distributional similarities, I am somewhat reluctant to label *a*-adjectives as PPs or locative particles. PPs are, of course, phrasal, while *a*-adjectives are single words. Locative particles are single words, but they are morphologically simplex while also appearing to form a closed list. By contrast, *a*-adjectives have a complex morphological structure and appear to be open-ended, as illustrated in 8. Ultimately, for the present study of language acquisition, the central issue is not what these *a*-adjectives are labeled by the theorist but how they are treated by the child learner on the basis of linguistic evidence.

In §4, I show that the child learner can make use of the distributional evidence reviewed here to establish the properties of *a*-adjectives. Before that, we turn to a detailed reassessment of Boyd and Goldberg's (2011) proposal of statistical preemption.

3. On statistical preemption. Boyd and Goldberg's approach has well-known precedents (e.g. the principle of uniqueness (Wexler & Culicover 1980), and the principle of contrast (Clark 1987)), which fall under the broad tenet of indirect negative evidence (Chomsky 1981; see Pinker 1989 for a review). In the present case, the learner would consider two alternative hypotheses for *a*-adjectives: one allows attributive/prenominal use (*P*) and the other does not (¬*P*). A child who consistently observes attested examples for ¬*P* (e.g. *the cat that is asleep*) may be led to conclude that the alternative hypothesis *P* is prohibited. In other words, direct positive evidence for ¬*P* constitutes indirect negative evidence against *P*. Presumably, this inclination will grow stronger as the evidence for ¬*P* accumulates; thus the preemptive effect ¬*P* over *P* would be statistical. In this section, the formal, empirical, and experimental aspects of the statistical preemption hypothesis are discussed.

3.1. Formal and empirical problems. The intuition behind statistical preemption seems clear, but there are a number of general as well as specific problems with this account. It is unclear how statistical preemption, or any kind of indirect negative evidence, is actually used in a psychological theory of learning.

To deploy indirect negative evidence, the learner needs to decide which hypothesis is underrepresented in the learning data. To do so in general requires comparing the extensions of the competing hypotheses, which is computationally prohibitive and may even be uncomputable (Osherson et al. 1986, Fodor & Sakas 2005). In the present case, the learner needs to calculate and compare the posterior probabilities of the hypotheses *P* and ¬*P* given the learning data presented so far. Boyd and Goldberg offer no suggestion as to how this may be done by the child learner. The most relevant statistical models of inference proposed elsewhere in the literature (Tenenbaum & Griffiths 2001, Chater & Vitányi 2007) treat learning in an abstract setting: for instance, the learner performs iterative optimization over the entire corpus of the input so the best hypothesis may be chosen. As such, these models are explicitly dissociated from psychological learning mechanisms.[1] In a typical formulation of the Bayesian learning framework, the learner may have undifferentiated a priori probabilities of *P* and ¬*P*; that is, both attributive and nonattributive usage are allowed for *a*-adjectives. The consistent absence of expressions such as *the asleep cat* will gradually lower the posterior probability of *P*, however, leaving ¬*P* (nonattributive usage) as the favored hypothesis. Thus, the Bayesian formulation is quite clearly a form of indirect negative evidence.

More importantly, and more empirically, statistical preemption makes a set of incorrect predictions about the outcome of learning when tested on a large sample of child-directed English. The assessment here draws from a parsed corpus of just over 180,000 child-directed English utterances, about 440,000 words in all (Pearl & Sprouse 2013), which facilitates the search for syntactic structures involving the usage patterns of adjectives. The parsed corpus is also supplemented by a 4.3-million-word text corpus of child-directed English. The parsed corpus contains twelve *a*-adjectives, and no additional *a*-adjective is found in the much larger text corpus.[2]

[1] For recent empirical studies, see Perfors et al. 2010 and Villavicencio et al. 2013.

[2] Instances of *alive* used attributively (*alive things walk*) by both children and adults were found, including *A thumping! A bumping! A wild alive scratching!* from the well-known Dr. Seuss book *Horton hatches the egg*. Presumably, the adjective *alive* is an exceptional member of the *a*-adjective class, which can be learned from attested usage just like the acquisition of English irregular verbs as exceptions to the 'add *-d*' rule. This word is excluded from further consideration.

(16) across, afraid, ahead, alike, alone, apart, around, ashamed, asleep, awake, aware, away

The frequencies of these *a*-adjectives are unremarkable. While they are all relatively frequent enough to occur in a modest 440,000-word sample of child-directed speech, none is frequent enough such that its absence from attributive usage would be conspicuous, which indeed has been suggested as another strategy for indirect negative evidence (e.g. Stefanowitsch 2008). In the parsed corpus, there are 517 predicatively used adjectives (e.g. *the cat is nice*), with an average frequency of 13.75, and 575 attributively used adjectives (e.g. *the big cat*), with an average frequency of 14.73. The intersection of the two sets consists of 198 adjectives that are used both predicatively and attributively, with an average frequency of 57.7, which provides more opportunities for them to be used in both constructions. But only one of the twelve *a*-adjectives (*afraid*, which appears seventy-three times) falls in this higher-frequency range, and many of the other eleven *a*-adjectives appear only once or twice. At the same time, even a cursory search reveals that the corpus contains other adjectives (e.g. *careful*) that are much more frequent than *afraid* but appear exclusively predicatively—but unlike the *a*-adjectives, these adjectives can appear attributively. Frequency, then, is not a reliable cue for the admissibility of syntactic usage.

To test Boyd and Goldberg's statistical preemption hypothesis more directly, a search was done for all adjectives appearing predicatively in relative clauses (e.g. *Do you have anything that is sharp?*). A total of sixty-five such relative clauses were found, involving forty-five unique adjectives. If these adjectives appear exclusively in relative clauses, then their attributive usage must be preempted, according to the statistical preemption hypothesis. Therefore, a search in the parsed corpus for attributive usages of these forty-five adjectives was done, but only twenty-eight of them were found in attributive position. The remaining seventeen adjectives, which were used exclusively in relative clauses and not in noun phrases, are given in 17.

(17) *afraid*, *alike*, annoying, careful, *gone*, handicapped, heavier, interested, jolly, *off*, *on*, shaped, supposed, sure, torn, washable, wrapped

Those in italics (five in all) are indeed adjectives that resist attributive use, including two *a*-adjectives, but none of the other twelve are, leading to a false positive rate of 71% (12/17). Meanwhile, of the twelve genuine *a*-adjectives in the 180,000-utterance corpus, ten are not identified by statistical preemption, with a false negative rate of 83% (10/12).

It should be pointed out that 180,000 child-directed utterances do not constitute a sufficiently large sample of input to the learner. I therefore turned next to a 4.3-million-word child-directed text corpus from the CHILDES database (MacWhinney 2000), which corresponds roughly to a year's worth of input data for some English children (see §4 below). The text corpus includes the data from which the parsed corpus (Pearl & Sprouse 2013) is derived. If statistical preemption works effectively, then the adjectives falsely identified as attributively resistant in 17 should appear in prenominal positions in the much larger database. Unfortunately, this is not the case: seven out of the twelve adjectives (*annoying*, *careful*, *interested*, *jolly*, *torn*, *washable*, *wrapped*) still fail to appear prenominally and remain false positives. Although the text corpus does not allow for efficient and accurate searching for relative clauses, we can be fairly certain that it will contain additional adjectives, like those in 17, that appear exclusively in relative clauses: these would be false positives under the statistical preemption hypothesis. Similarly, the *a*-adjectives in 16 were searched for in the 4.3-million-word text corpus. I wanted to see how many of these would be used in a relative clause such that their attributive use could be preempted. Only one more (*asleep*) joins the ranks of the

two identified in 17, while the remaining nine remain unaccounted for and would be classified as false negatives.

This corpus analysis also suggests that, at least in the present case study, the probabilistic aspect of statistical preemption is unlikely to play any role in language learning. As the search results show, adjectives are very rarely used predicatively in relative clauses: only sixty-five occur in over 180,000 utterances (or in 0.0003% of the utterances), and these are further spread over forty-five unique adjectives. In other words, if an adjective is used in a relative clause at all, it is very likely used only once. The child will not encounter cumulative evidence for $\neg P$ over multiple attestations so as to gradually weaken the alternative hypothesis P.

Finally, the statistical preemption hypothesis makes an additional prediction for TYPICAL adjectives that, as far as I can tell, is also incorrect. If the ungrammaticality of the attributive use (P) is due to the mutually exclusive blocking effect of their paraphrastic alternatives ($\neg P$), then the relative clause use of an adjective ($\neg P$) ought to be blocked if the learner witnesses only its prenominal usage (P). The very low-frequency usage of adjectives in relative clauses in the corpus search results suggests that the vast majority of adjectives appear exclusively attributively when modifying noun phrases. Indeed, numerous adjectives in the corpus, ranging from very frequent ones (e.g. *green*, which appears thousands of times) to relatively rare ones (e.g. *ancient*), pattern this way, and therefore should be prevented from appearing as predicates in relative clauses altogether according to the logic of statistical preemption. But this is clearly incorrect.

Taken together, it is very unlikely for the child to receive sufficient evidence for preemption, statistically or otherwise, of the attributive use of *a*-adjectives. Furthermore, statistical preemption would mislead the learner to falsely identify a small but nonnegligible number of adjectives as attributively resistant (see 17) and to identify a very large number of adjectives as inadmissible in relative clauses.

3.2. *A*-ADJECTIVES IN EXPERIMENTAL STUDIES. We now consider Boyd and Goldberg's (2011) three experimental studies in order to reassess the role of statistical preemption. In all three experiments, adult English speakers consistently and strongly resisted the attributive use of existing *a*-adjectives, confirming the restriction observed in previous research (2). The crucial experiments created novel *a*-adjective-like adjectives (e.g. *ablim*). The subjects were exposed to the various usages of these adjectives so that it could be determined whether their attributive usage in a natural production task was affected as a result.

Experiment 1 established a baseline performance. Without exposure to any usage examples, subjects were significantly less likely to use novel *a*-adjectives (e.g. *ablim*) than novel non-*a*-adjectives (e.g. *chammy*) in attributive position. Boyd and Goldberg take this result to mean that the subjects have formed a category for *a*-adjective-like adjectives that resist attributive use, although they do not provide specific criteria by which such a category is formed. Experiment 2 is the most critical study, as it investigates the preemptive effect of relative clause usage (e.g. *the pig that was ablim moved to the star*). During the exposure phase, some of the novel *a*-adjectives appeared in relative clauses while others did not. Nevertheless, all of the novel *a*-adjectives showed considerably lower attributive usage rates than novel non-*a*-adjectives and in fact were at a level similar to that of the existing *a*-adjectives. Boyd and Goldberg conclude that relative clause usage preempted the attributive form while also allowing the subjects to generalize over the entire class of novel *a*-adjectives. Experiment 3 involved exposure patterns such as *the hamster that's ablim and proud of himself moved to the star*: subjects in these conditions appeared to treat *ablim* more like a typical adjective in allow-

ing for attributive use. Boyd and Goldberg suggest that the additional modification (*proud of himself*), which independently prohibits attributive use, caused the subjects to discount the preemptive effect from the relative clause.

There is an alternative account of these findings that Boyd and Goldberg overlooked, one that is also consistent with the distributional analysis of *a*-adjectives reviewed in §2. That is, the subjects in Boyd and Goldberg's experiments could simply be extending their syntactic and morphological knowledge as native speakers of English to the novel *a*-adjectives. While none of the created *a*-adjectives (*ablim*, *adax*, *afraz*, and *agask*) contains an actual word or stem, the syllable that remains after the segmentation of the initial schwa *a-* follows the phonotactics of English and could be interpreted as a potential (and unknown) word describing some property of the animals used in the experiments.[3] This is especially likely because the subjects were explicitly instructed prior to the experiment that they would encounter words they had never seen before and should produce descriptions of the situation like a native English speaker (Boyd & Goldberg 2011:67). Additionally, *a*-adjectives are relatively common words. On average they are three times more frequent than schwa (*a-*) initial non-*a*-adjectives such as *above* and *alert*, as in 8. This also favors the analysis of *ablim* as an *a*-adjective upon its presentation. In other words, the subjects would treat the novel *a*-adjectives much as they do words like *wugs* and *ricked* in Berko's classic study (1958) by extending their native knowledge of English morphology. This immediately accounts for the results in experiment 1: even without exposure to any usage patterns for the novel *a*-adjectives, subjects spontaneously limit their attributive use to a level significantly below that of the novel non-*a*-adjectives. Experiment 2 presented the subjects with the novel *a*-adjectives in relative clauses. This strengthens the morphologically based analysis (experiment 1) so that subjects would be even more certain that words such as *ablim* behave like *a*-adjectives and would further avoid attributive use. Furthermore, the use of relative clauses in the exposure phase encourages relative clause usage in the subjects' productions, due to the well-established priming effects of syntactic structures (e.g. Bock 1986, Pickering & Branigan 1998), thereby reducing the rate of attributive use even more, which is exactly what experiment 2 shows. Thus, the main results of Boyd and Goldberg's studies are consistent with the observation (Salkoff 1983, Larson & Marušič 2004, Coppock 2008, etc.) that speakers of English are aware of the morphological and syntactic restrictions on *a*-adjectives, and these restriction can thus be extended to novel items; see 8. This, however, raises the question of how English speakers acquire these properties of *a*-adjectives in the first place.

Regarding experiment 3, my interpretation also differs from Boyd and Goldberg's analysis. It seems likely that the subjects took the conjoined structure *ablim and proud of himself* as evidence that *ablim* patterns exactly like *proud*, and that the additional modification of *of himself* is superfluous. Although different syntactic categories can be conjoined (Sag et al. 1985), English usage data overwhelmingly favors a uniform treatment of conjoined elements. For instance, in the 180,000 parsed child-directed English utterances (Pearl & Sprouse 2013), there are 348 conjoined adjective phrases, all of which involve two or more adjectives of the same type (e.g. *soft or hard*; *red, blue, and white*). In any case, it is highly unlikely that the materials in experiment 3 will have any effect in actual language acquisition. As noted earlier in this section, the frequency of these particular relative clauses is already extremely low (occurring in sixty-five out of 180,000 utterances), and only one involved conjoined adjectives (*those are special kinds of hairs*

[3] I thank a referee for raising this possibility.

that are sharp and pointy). The likelihood of finding an *a*-adjective in a relative clause that is conjoined with another adjective and followed by yet another modification that independently avoids attributive use—for example, *ablim and proud of himself* as in experiment 3—is vanishingly low in any realistic sample of child-directed speech.

In sum, the statistical preemption approach makes incorrect empirical predictions about the outcome of learning. Boyd and Goldberg's experimental results appear only to confirm the speaker's syntactic and morphological knowledge about *a*-adjectives, which readily extends to novel items. The problem of how such knowledge is acquired thus remains unanswered, and the absence of any quantitative analysis of realistic input data further undermines the feasibility of statistical preemption. I now show that positive evidence for the grammatical properties of *a*-adjectives (§2) is robustly attested in child-directed English to guide language acquisition.

4. GENERALIZATION WITH POSITIVE EVIDENCE. The general strategy for learning negative constraints can be outlined as follows. The learner, on the basis of positive evidence, observes that *a*-adjectives pattern distributionally like locative particles and prepositional phrases; thus the resistance to attributive usage in these two classes of linguistic units can be extended to the *a*-adjectives. However, not all *a*-adjectives are used in a PP-like context in child-directed input; an independently motivated principle of generalization is necessary for the learner to extend the property of attributive resistance to the entire class of *a*-adjectives.

The empirical study here focuses on the distributional properties of *a*-adjectives in a 4.3-million-word corpus of child-directed speech from the CHILDES database (MacWhinney 2000).[4] While more data would be desirable, this is not a trivial sample: it corresponds to more than a year of child-directed speech for children at the lowest socioeconomic level, more than a half of a year for children from working-class families, and more than a third of a year for children from professional families (Hart & Risley 2003). To the extent that all speakers seem to learn the attributive prohibition on *a*-adjectives, the corpus provides a representative sample of the language acquisition data.

Boyd and Goldberg (2011) did not provide any analysis of child language, nor am I aware of any previous acquisition studies of *a*-adjectives. To address this issue, I searched through an approximately 1.9-million-word corpus of child English from the CHILDES database (MacWhinney 2000) and extracted the usage patterns of *a*-adjectives, with a total of about 2,300 tokens. Not a single instance of attributive usage of an *a*-adjective was found (though see n. 2). Although the child data is from a large number of subjects and recording sessions, the average age of the learners is just over 2;10. This suggests that English-learning children acquire the syntactic properties of *a*-adjectives very early, and the distributional evidence for their acquisition must be robustly available.

4.1. POSITIVE EVIDENCE. The attested *a*-adjectives were first extracted from the child-directed input. In order to identify the morphological structure of *a*-adjectives, the learner needs to discover the compositional formation of *a-* with a stem. Morpheme segmentation takes place at a very early age (Gerken & McIntosh 1993), and crosslinguistically,

[4] The data draws from the American English portion of the CHILDES database, which can be found at http://childes.psy.cmu.edu/data/Eng-NA-MOR/. Although the CHILDES project recommends a consistent coding format for children's speech, with each utterance marked as '*CHI', not all transcripts follow this convention, making it impossible to automatically distinguish child from child-directed speech in those files. Thus, both child and child-directed speech were extracted from only those transcription files that marked children's utterances with '*CHI', which nevertheless includes the vast majority of the American English acquisition data in the public domain (2,133 out of the 2,279 transcript files in the CHILDES database).

children's morphological knowledge is generally highly reliable (Guasti 2004). English-learner's word-segmentation errors (Brown 1973, Peters 1983, Yang 2004) such as *there are three dults in our family* (*dult* from *a-dult*) and *I was have* (*have* from *be-have*) further suggest that children recognize affix-like elements as phonological units during lexical acquisition. The morphological segmentation of *a*-adjectives would be facilitated by the fact that *a*-adjective stems are highly frequent and thus very likely to be part of a young child's vocabulary. For our quantitative analysis, the corpus data was processed with a part-of-speech tagger to extract all of the adjectives. Word-initial unstressed schwa *a-* was then segmented off without violating the phonotactics of English; the results are presented in 18.

(18) a. Containing stems: afraid, awake, aware, ashamed, ahead, alone, apart, around, asleep, alike, away, across
b. Not containing stems: amazing, annoying, allergic, available, adorable, another, American, attractive, approachable, acceptable, agreeable, affectionate, adept, above, aberrant

Three distributional patterns emerge from the child-directed corpus. First, the presence or absence of a stem partitions the adjectives into two distinct classes. All of the items in 18a contain a stem, and all are *a*-adjectives. By contrast, none of the items in 18b contains a stem—and none are *a*-adjectives and all can be used attributively. Thus, the morphological criterion defines membership for the *a*-adjectives well.

Second, I searched for *right*-type modification usage for the non-*a*-adjectives in 18b.[5] Not a single example like those in 13 was found, which supports the use of *right*-type modification as a diagnostic to distinguish *a*-adjectives from non-*a*-adjectives.

Third, the compositionally formed *a*-adjectives in 18a show robust usage with *right*-type modification in eight out of the twelve types. A corpus example for each *a*-adjective is provided in 19.

(19) a. are you wide awake?
b. I'm well aware of my shortcomings thank you
c. go right ahead.
d. it fell right apart on you.
e. turn right around.
f. finish the book right away.
g. he fell fast asleep.
h. we are coming right across.

The number of attested examples ranges from a handful (for *apart*) to over a hundred (for *away*). While the 4.3-million-word corpus used here is not trivial in size, positive evidence for the distributional properties of *a*-adjectives will be even more robust in larger samples of the primary linguistic data. I further note that the corpus contains numerous instances of *right*-type modification with locative particles (14) as well as prepositional phrases (15): *right here* and *right there* appear over 3,000 times, *right up*/*over*/*on* are in the hundreds, *right off*/*down*/*under* dozens, and so forth. Therefore, a typical English-learning child will have plenty of opportunities to observe that *a*-adjectives, locative participles, and prepositional phrases are distributionally similar on the basis of positive evidence.

We are not quite done. The learner still needs to form generalizations about the *a*-adjectives as a class: after all, only eight of the twelve members of the *a*-adjectives in

[5] I thank a referee for this suggestion.

the input (18a) are used with *right*-type modification. But that is the TYPICAL situation in language acquisition. In almost all cases of language learning, the child will not be able to witness the entire range of syntactic behavior for every member of a linguistic class. Case in point: there are forty *a*-adjectives in 9 that prohibit attributive use. Yet only twelve are attested in roughly a year's worth of child-directed English, and only twenty-eight appear at all in a 51-million-word spoken American English corpus (Brysbaert & New 2009). We can also be fairly certain that not all twenty-eight, perhaps only a minority of them, will be used with *right*-type modification, the signature distributional evidence that relates *a*-adjectives to locative particles and prepositional phrases. To the extent that English speakers generally resist the attributive use of *a*-adjectives, even with novel words that have a similar morphological structure (8) and in Boyd and Goldberg's experiments, this knowledge must be acquired on a small but highly frequent set of items available in child-directed English.

Thus, a productive generalization over a class of lexical items should, and in fact must, be acquired if the learner witnesses SUFFICIENT positive evidence over its members. (Conversely, if learners do not witness enough positive instances, they will decide the generalization is unproductive, proceed to lexicalize the positively attested examples, and refrain from extending the pattern to novel items.) The key question, then, is what counts as SUFFICIENT positive evidence. Again, this is the typical question in language acquisition. To take a well-known example, English children learn that the *-ed* rule is productive presumably because they observe a sufficiently large number of regular verbs following the rule, despite the presence of some 120 irregular exceptions (Berko 1958, Marcus et al. 1992): the child will not, and should not need to, encounter the past-tense form of every regular verb.

4.2. GENERALIZATION. Let us consider the nature of generalization with a toy example from a nonlinguistic domain. If I came across ten species of birds with red crowns, eight of which are tame, I would probably let my guard down and assume the next one I encounter to also be harmless: eight out of ten seems a pretty good batting average, even though I have had no direct experience with the remaining two. But one out of ten? Two out of ten? Without any information about the remaining majority, I would be well advised to proceed with caution: I have not had negative evidence that they would be hostile, but I have not seen sufficient positive evidence for their friendliness, either. In any problem of learning, a generalization needs to be supported by the weight of evidence.

Yang 2005 developed a formal learning model dubbed the TOLERANCE PRINCIPLE, which provides a sufficiency measure for generalization.

(20) TOLERANCE PRINCIPLE: If R is a productive rule applicable to N candidates, then the following relation holds between N and e, the number of exceptions that could but do not follow R:

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

The empirical motivation for the tolerance principle comes from psycholinguistic research. There is evidence that lexical exceptions slow down the real-time processing of items that follow rules (Swinney & Cutler 1979, Penke & Krause 2002, Fleischhauer & Clahsen 2012, etc.). This allows for a cost/benefit calculus of processing complexity, with which the tolerable number of exceptions for a productive rule can be derived mathematically; see Yang 2016 for detailed discussion and many empirical case studies.

The application of the tolerance principle is illustrated with two simple examples. First, consider the acquisition of the English past tense. If a typical English speaker

knows $e = 120$ irregular verbs, the productivity of the *-ed* rule is guaranteed only if there are many more regular verbs. Specifically, there must be N verbs, including both regulars and irregulars, such that $\theta_N = N/\ln N \geq 120$. The minimum value of N is 800. In other words, if there are at least 680 regular verbs in English, the *-ed* rule can tolerate 120 irregular verbs. Since there are clearly more than 680 regular verbs in English, the learner will be justified to conclude that the *-ed* rule is productive and can be extended to novel items (Berko 1958).

Consider then an example from syntax. In a well-known article, Newmeyer (2004) regards exceptions as a serious challenge to the theory of parameters. For instance, while French generally places the adjective after the noun (*un livre noir* 'black book'), there is a special class of adjectives that precede the noun.

(21) a. une **nouvelle** maison 'a new house'
b. une **vieille** amie 'a friend for a long time'

Note that the problem of exceptions in 21 needs to be resolved under any theory of learning, and the assumption of parameters is not essential. For instance, in a nonparametric theory that uses rules or constructions, the learner still needs to recognize that, despite a small number of exceptional adjectives that appear before the noun, the general pattern is to place the adjective after the noun. To this end, a corpus of approximately 200,000 child-directed French sentences in the CHILDES database (MacWhinney 2000) was analyzed. The data was processed with a part-of-speech tagger to extract adjacent adjective and noun pairs, which were then subject to manual inspection. All adjectives were lemmatized. In all, $e = 20$ exceptional adjectives appear in a prenominal position: if the postnominal order is to be productive, there need to be approximately eighty typical adjectives to counterbalance the examples ($\theta_{100} = 100/\ln 100 = 22 > 20$). Indeed, by about halfway into the corpus, at least 120 unique postnominal adjectives had been located. A larger corpus can only increase this count, and the exceptional prenominal adjectives belong to a closed list. Thus, the French child will have ample evidence to the support the noun-adjective order, despite a tolerable number of exceptions.

The application of the tolerance principle to the present case is straightforward (see Yang 2016:Ch. 6 for details). If the learner encounters a set of N lexical items, a subset of at least $N - \theta_N$ items following a rule is sufficient for extending the rule over the entire set of N items. Here the learner encounters $N = 12$ attested *a*-adjectives, which are composed of the prefix *a-* and a stem. Out of these, the majority (eight out of twelve) show explicit distributional evidence of patterning with locative particles. The learner has not encountered any explicit evidence concerning the remaining four, but eight is in fact sufficient for generalization, as $N - \theta_N = 12 - 5 = 7$. This ensures that the property associated with locative particles, including the prohibition on attributive use, can productively hold for the entire class of *a*-adjectives. In other words, the four *a*-adjectives will inherit the property of attributive resistance from the eight members with positive evidence, just as the child can extend the *-ed* suffix from the regular verbs attested in the past tense to verbs that have never appeared in past tense, once the rule has been established as productive. We conclude that from a reasonable sample of child-directed English, the learner can acquire the properties of the *a*-adjectives and readily extend them to novel items as in 8, as well as in Boyd and Goldberg's (2011) experimental studies.

5. SUMMARY. In this article, I have shown that negative knowledge can be acquired from positive evidence, as long as the quantity of positive evidence is sufficient to facilitate distributional learning. Specifically, child-directed English input has been used to argue that *a*-adjectives can be reliably identified with locative particles and locative phrases, under

an independently motivated principle of generalization (Yang 2005, 2016). The strategy of statistical preemption (Boyd & Goldberg 2011) is neither sufficient nor necessary.

I emphasize that the learnability approach developed here is a most conservative kind. It is possible that the child has access to other sources of information that provide more direct constraints against the prenominal use of the *a*-adjectives. For instance, the syntax and semantics of *a*-adjectives require additional research, and a deeper understanding of them may contribute to solving the problem of child language acquisition. If so, the learner's task would presumably be further simplified. Indeed, my proposal deals with a worst-case situation, in which the child is to acquire the properties of the *a*-adjectives with the most basic and traditional use of distributional learning across contexts: if it walks like a duck and quacks like a duck, it must be a duck (or a locative particle). The use of the tolerance principle is also programmatic: SOME model of generalization in the face of exceptions must be at work for language acquisition. In the present case, at least, a simple majority vote (eight out of twelve) will suffice to derive the same learning outcome.

I conclude with some brief remarks on the use of corpora in the study of language acquisition. The availability of electronic databases in recent years has provided linguists with an ever-expanding and readily accessible source of data. For instance, previous researchers (Boyd & Goldberg 2011, Bruening 2011b, Goldberg 2011) have made use of the Corpus of Contemporary American English (Davies 2008) as well as Web examples in the analysis of *a*-adjectives. But the primary linguistic data for child language acquisition may be quite different from the type found in large-scale corpora. As we have seen, the hypothesis of statistical preemption requires attestations of paraphrastic alternatives: while these forms can be constructed by the theorists or administered in experimental studies, hardly any are found in realistic language-learning data. In a similar vein, although ungrammatical examples can be constructed by the theorist to further the study of language, they are of little relevance for the acquisition of language as the child does not have access to such grammatically informative data.

It is widely known that only a small fraction of the linguistically possible forms are attested in language use (Zipf 1949, Jelinek 1998). This poses interesting challenges for the study of language acquisition. The low diversity of linguistic forms in child language may lead to underestimation of children's grammatical capabilities (e.g. Tomasello 2000), while assessments that take the statistical properties of language into consideration can reach opposite conclusions (e.g. Valian et al. 2009, Yang 2013). At the same time, the likewise skewed distribution in child-directed input suggests that it is implausible for the learner to witness all, or even a majority, of syntactic forms that can be generated by the target grammar: negative evidence is surely absent in language acquisition, but even positive evidence will only be sparsely represented. By necessity, the properties of the *a*-adjectives as a class can only be acquired on the basis of very few members such as those in 13a. The decisions in language acquisition therefore will likely rest on the behavior of a small but highly frequent set of lexical items; see Yang 2016:Ch. 7 for extended discussion. Children must be equipped to connect seemingly disparate phenomena, and they will need to generalize aggressively from early on.

REFERENCES

BAKER, CARL L. 1979. Syntactic theory and the projection problem. *Linguistic Inquiry* 10.533–81. Online: http://www.jstor.org/stable/4178133.

BEARD, ROBERT. 1995. *Lexeme-morpheme base morphology*. Albany: SUNY Press.

BERKO, JEAN. 1958. The child's learning of English morphology. *Word* 14.150–77.

BERWICK, ROBERT. 1985. *The acquisition of syntactic knowledge*. Cambridge, MA: MIT Press.
BOCK, J. KATHRYN. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18.355–87. DOI: 10.1016/0010-0285(86)90004-6.
BOLINGER, DWIGHT. 1971. *The phrasal verb in English*. Cambridge, MA: Harvard University Press.
BOULDIN, JOHN MICHAEL. 1990. *The syntax and semantics of postnominal adjectives in English*. Minneapolis: University of Minnesota dissertation.
BOWERMAN, MELISSA. 1988. The 'no negative evidence' problem: How do children avoid constructing an overly general grammar? *Explaining language universals*, ed. by John A. Hawkins, 73–101. Oxford: Blackwell.
BOYD, JEREMY K., and ADELE E. GOLDBERG. 2011. Learning what NOT to say: The role of statistical preemption and categorization in *a*-adjective production. *Language* 87.55–83. DOI: 10.1353/lan.2011.0012.
BRAINE, MARTIN D. S. 1971. On two types of models of the internalization of grammars. *The ontogenesis of grammar: A theoretical symposium*, ed. by Daniel I. Slobin, 153–86. New York: Academic Press.
BROWN, ROGER. 1973. *A first language: The early stages*. Cambridge: Harvard University Press.
BROWN, ROGER, and CAMILLE HANLON. 1970. Derivational complexity and the order of acquisition in child speech. *Cognition and the development of language*, ed. by John R. Hayes, 11–53. New York: Wiley.
BRUENING, BENJAMIN. 2011a. A-adjectives are PPs, not adjectives. Newark: University of Delaware. Blogpost. Online: http://lingcomm.blogspot.com/2011/03/adjectives-are-pps-not-adjectives.html.
BRUENING, BENJAMIN. 2011b. A-adjectives again: Response to Goldberg. Newark: University of Delaware. Blogpost. Online: http://lingcomm.blogspot.com/search?q=A-adjectives+again:+Response+to+Goldberg.
BRYSBAERT, MARC, and BORIS NEW. 2009. Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods* 41.977–90. DOI: 10.3758/BRM.41.4.977.
CAMPBELL, AIMEE L., and MICHAEL TOMASELLO. 2001. The acquisition of English dative constructions. *Applied Psycholinguistics* 22.253–67.
CHATER, NICK, and PAUL VITÁNYI. 2007. 'Ideal learning' of natural language: Positive results about learning from positive evidence. *Journal of Mathematical Psychology* 51.135–63. DOI: 10.1016/j.jmp.2006.10.002.
CHOMSKY, NOAM. 1981. *Lectures in government and binding*. Dordrecht: Foris.
CINQUE, GUGLIELMO. 2010. *The syntax of adjectives: A comparative study*. Cambridge, MA: MIT Press.
CLARK, EVE V. 1987. The principle of contrast: A constraint on language acquisition. *Mechanisms of language acquisition*, ed. by Brian MacWhinney, 1–33. Hillsdale, NJ: Lawrence Erlbaum.
CONWELL, ERIN, and KATHERINE DEMUTH. 2007. Early syntactic productivity: Evidence from dative shift. *Cognition* 103.163–79. DOI: 10.1016/j.cognition.2006.03.003.
COPPOCK, ELIZABETH. 2008. *The logical and empirical foundations of Baker's paradox*. Stanford, CA: Stanford University dissertation.
DAVIES, MARK. 2008. *Corpus of Contemporary American English (COCA): 410+ million words, 1990–present*. Provo, UT: Brigham Young University. Online: http://corpus.byu.edu/coca/.
FLEISCHHAUER, ELISABETH, and HARALD CLAHSEN. 2012. Generating inflected word forms in real time: Evaluating the role of age, frequency, and working memory. *Proceedings of the Boston University Conference on Language Development (BUCLD)* 36.1.164–76.
FODOR, JANET DEAN. 1989. Learning the periphery. *Learnability and linguistic theory*, ed. by Robert J. Matthews and William Demopoulos, 129–54. Berlin: Springer.
FODOR, JANET DEAN, and STEPHEN CRAIN. 1987. Simplicity and generality of rules in language acquisition. *Mechanisms of language acquisition*, ed. by Brian MacWhinney, 35–63. Mahwah, NJ: Lawrence Erlbaum.

FODOR, JANET DEAN, and WILLIAM GREGORY SAKAS. 2005. The subset principle in syntax: Costs of compliance. *Journal of Linguistics* 41.513–69. DOI: 10.1017/S002222670 5003439.

GERKEN, LOUANN, and BONNIE J. MCINTOSH. 1993. Interplay of function morphemes and prosody in early language. *Developmental Psychology* 29.448–57. DOI: 10.1037/0012 1649.29.3.448.

GOLDBERG, ADELE E. 2011. Are a-adjectives like *afraid* prepositional phrases underlying and does it matter from a learnability perspective? Princeton, NJ: Princeton University, MS. Online: https://www.princeton.edu/~adele/Media/Pubs-by-topic_files/a-adjsnotpps .pdf.

GUASTI, MARIA TERESA. 2004. *Language acquisition: The growth of grammar*. Cambridge, MA: MIT Press.

HART, BETTY, and TODD R. RISLEY. 2003. The early catastrophe: The 30 million word gap by age 3. *American Educator* 27.4–9. Online: http://www.aft.org/ae/spring2003/hart _risley.

HUDDLESTON, RODNEY, and GEOFFREY K. PULLUM. 2001. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.

INAGAKI, SHUNJI. 1997. Japanese and Chinese learners' acquisition of the narrow-range rules for the dative alternation in English. *Language Learning* 47.637–69. DOI: 10.1111/0023 -8333.00024.

JELINEK, FREDERICK. 1998. *Statistical methods for speech recognition*. Cambridge, MA: MIT Press.

LARSON, RICHARD K., and FRANC MARUŠIČ. 2004. On indefinite pronoun structures with APs: Reply to Kishimoto. *Linguistic Inquiry* 35.268–87. DOI: 10.1162/002438904323 019075.

LONG, RALPH B. 1969. *The sentence and its parts*. Chicago: University of Chicago Press.

MACWHINNEY, BRIAN (ed.) 1987. *The mechanisms of language acquisition*. New York: Psychology Press.

MACWHINNEY, BRIAN. 2000. *The CHILDES project: Tools for analyzing talk*. 3rd edn. Mahwah, NJ: Lawrence Erlbaum.

MARCUS, GARY F. 1993. Negative evidence in language acquisition. *Cognition* 46.53–85. DOI: 10.1016/0010-0277(93)90022-N.

MARCUS, GARY F.; STEVEN PINKER; MICHAEL T. ULLMAN; MICHELLE HOLLANDER; JOHN ROSEN; and FEI XU. 1992. *Overregularization in language acquisition*. (Monographs of the Society for Research in Child Development.) Chicago: University of Chicago Press.

MAZURKEWICH, IRENE, and LYDIA WHITE. 1984. The acquisition of the dative alternation: Unlearning overgeneralizations. *Cognition* 16.261–83. DOI: 10.1016/0010-0277(84) 90030-1.

NEWMEYER, FREDERICK J. 2004. Against a parameter-setting approach to typological variation. *Linguistic Variation Yearbook* 4.181–234. DOI: 10.1075/livy.4.06new.

OSHERSON, DANIEL N.; MICHAEL STOB; and SCOTT WEINSTEIN. 1986. *Systems that learn: An introduction to learning theory for cognitive and computer scientists*. Cambridge, MA: MIT Press.

PEARL, LISA, and JON SPROUSE. 2013. Syntactic islands and learning biases: Combining experimental syntax and computational modeling to investigate the language acquisition problem. *Language Acquisition* 20.23–68. DOI: 10.1080/10489223.2012.738742.

PENKE, MARTINA, and MARION KRAUSE. 2002. German noun plurals: A challenge to the dual-mechanism model. *Brain and Language* 81.303–11. DOI: 10.1006/brln.2001 .2526.

PERFORS, AMY; JOSHUA B. TENENBAUM; and ELIZABETH WONNACOTT. 2010. Variability, negative evidence, and the acquisition of verb argument constructions. *Journal of Child Language* 37.607–42. DOI: 10.1017/S0305000910000012.

PESETSKY, DAVID. 1995. *Zero syntax: Experiencer and cascade*. Cambridge, MA: MIT Press.

PETERS, ANN M. 1983. *The units of language acquisition*. Cambridge: Cambridge University Press.

PICKERING, MARTIN J., and HOLLY P. BRANIGAN. 1998. The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language* 39.633–51. DOI: 10.1006/jmla.1998.2592.

PINKER, STEVEN. 1989. *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press.
POSER, WILLIAM J. 1992. Blocking of phrasal constructions by lexical items. *Lexical matters*, ed. by Ivan Sag and Anna Szabolsci, 111–30. Stanford, CA: CSLI Publications.
RANDALL, JANET H. 1990. Catapults and pendulums: The mechanics of language acquisition. *Linguistics* 28.1381–1406. DOI: 10.1515/ling.1990.28.6.1381.
RAUH, GISA. 1993. On the grammar of lexical and non-lexical prepositions in English. *The semantics of prepositions*, ed. by Cornelia Zelinsky-Wibbelt, 99–150. Berlin: Walter de Gruyter.
SAG, IVAN A.; GERALD GAZDAR; THOMAS WASOW; and STEVEN WEISLER. 1985. Coordination and how to distinguish categories. *Natural Language and Linguistic Theory* 3.117–71. DOI: 10.1007/BF00133839.
SALKOFF, MORRIS. 1983. Bees are swarming in the garden: A systematic synchronic study of productivity. *Language* 59.288–346. DOI: 10.2307/413576.
STEFANOWITSCH, ANATOL. 2008. Negative entrenchment: A usage-based approach to negative evidence. *Cognitive Linguistics* 19.513–31. DOI: 10.1515/COGL.2008.020.
STVAN, LAUREL SMITH. 1998. *The semantics and pragmatics of bare singular noun phrases*. Evanston, IL: Northwestern University dissertation.
SWINNEY, DAVID A., and ANNE CUTLER. 1979. The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior* 18.523–34. DOI: 10.1016/S0022-5371(79)90284-6.
TENENBAUM, JOSHUA B., and THOMAS L. GRIFFITHS. 2001. Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences* 24.629–40. DOI: 10.1017/S0140525X01000061.
TOMASELLO, MICHAEL. 2000. Do young children have adult syntactic competence? *Cognition* 74.209–53. DOI: 10.1016/S0010-0277(99)00069-4.
VALIAN, VIRGINIA; STEPHANIE SOLT; and JOHN STEWART. 2009. Abstract categories or limited-scope formulae? The case of children's determiners. *Journal of Child Language* 36.743–78. DOI: 10.1017/S0305000908009082.
VILLAVICENCIO, ALINE; MARCO IDIART; ROBERT C. BERWICK; and IGOR MALIOUTOV. 2013. Language acquisition and probabilistic models: Keeping it simple. *Proceedings of the 51st annual meeting of the Association for Computational Linguistics*, 1321–30. Online: http://www.aclweb.org/anthology/P13-1130.
WEXLER, KENNETH, and PETER CULICOVER. 1980. *Formal principles of language acquisition*. Cambridge, MA: MIT Press.
YANG, CHARLES. 2004. Universal grammar, statistics or both? *Trends in Cognitive Sciences* 8.451–56. DOI: 10.1016/j.tics.2004.08.006.
YANG, CHARLES. 2005. On productivity. *Linguistic Variation Yearbook* 5.333–70. DOI: 10.1075/livy.5.09yan.
YANG, CHARLES. 2013. Ontogeny and phylogeny of language. *Proceedings of the National Academy of Sciences* 110.6324–27. DOI: 10.1073/pnas.1216803110.
YANG, CHARLES. 2016. *Price of productivity*. Cambridge, MA: MIT Press, to appear.
ZIPF, GEORGE KINGSLEY. 1949. *Human behavior and the principle of least effort: An introduction to human ecology*. Cambridge, MA: Addison-Wesley.

Department of Linguistics and
Computer and Information Science
Institute for Research in Cognitive Science
619 Williams Hall
University of Pennsylvania
Philadelphia, PA 19081
[charles.yang@ling.upenn.edu]

[Received 22 May 2014;
revision invited 25 September 2014;
revision received 26 October 2014;
accepted 1 March 2015]