Negative Knowledge from Positive Evidence

Charles Yang

Department of Linguistics & Computer and Information Science

Institute for Research in Cognitive Science

619 Williams Hall

University of Pennsylvania

Philadelphia, PA 19081

charles.yang@ling.upenn.edu

(To appear in *Language*)

Negative Knowledge from Positive Evidence

Abstract

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 in language acquisition research is 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). In this article, we provide 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 productively extended to the adjectives in questions.

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 for 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, for 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 (1979): How does the child delineate the possible from the impossible based only on positive linguistic examples? Consider the dative constructions in English:

(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 does the child know that the double object construction is ungrammatical for verbs such as *donate* (1b) while they encounter plenty of positive instances of interchangeability between the two dative constructions, as in (1a)? The absence of negative evidence to the child (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, MacWhinney 1987, Fodor & Crain 1987, Pinker 1989, Fodor 1989, Randall 1990, Pesetsky 1995, Inagaki 1997, Campbell & Tomasello 2001, Conwell & Demuth 2007, etc.). In this paper, we 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, or the theorist, into the logical and empirical quagmire of Baker's Paradox. Negative knowledge of language can be acquired on the basis of positive evidence alone, as we illustrate with the case of the so-called a-adjectives.

It has been long observed (Bolinger 1971, Bouldin 1990, Huddleston & Pullum 2001, Larson & Marušič 2004, Cinque 2010) that a class of English adjectives 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 paper, Boyd & 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 paper, we consider an alternative learning strategy that makes use of positive evidence to learn negative exceptions. Following and extending analyses put forward by Salkoff (1983), Larson & Marušič (2004), Coppock (2008), Bruening (2011a,b) and others, we first present morphological and syntactic evidence that a-adjectives pattern with locative particles such as *up*, *out*, *over*, *here*, *there* etc and prepositional phrases. Such distributional equivalence enables the learner to extend the prohibition on attributive usage from locative particles and prepositional phrases to a-adjectives. In section 3, we discuss the formal and empirical problems with the statistical preemption approach, while offering an alternative interpretation of Boyd & Goldberg (2011)'s experimental results. In section 4, we provide quantitative analyses of a 4.3 million word corpus of child directed English: the learner has sufficient evidence to establish the distributional properties of a-adjectives under an independently motivated learning principle for generalization (Yang 2005). We conclude with a discussion of the necessity and challenges with the use of corpus data in the study of child language acquisition.

2   A-adjectives are not Atypical

Historically, many of the a-adjectives were derived from prepositional phrases (Long 1969). 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 distributional profile.

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 prepositional phrases behave like locative particles:

(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 (Huddleston & Pullum 2001), and the same holds for locative particles and prepositional phrases:

(5)   a.   ? the nocturnally awake cat
           ? a frequently away parent
           ? an aware, amused look (Huddleston & Pullum 2001, p559)

       b.   ? the almost here train
           ? the never present advisor
           ? an up and down experience

       c.   ? the always on time teacher
           ? an angry but on fire hitter
           ? a still in service shuttle

As noted by Salkoff (1983, p299) and Coppock (2008, p181), 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 the a-adjectives per se but with the aspectual prefix *a-*, which may be combined with and extended to novel words:

(6)   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 & Marušič (2004) observe that all a-adjectives can be decomposed into the prefix *a-* and a stem which 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 below is taken from their paper (p270) with a few of my own additions; none is generally acceptable in attributive use.

(7)   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., non-prefix) and a non-stem (8a) or semantically opaque stem (8b):

(8)   a.   The above examples
           The aloof professor
           The alert student
           The astute investor

       b.   The acute problem

Of course, the language acquisition problem does not go away under the morphological approach to a-adjectives. First, the learner needs to recognize that the *a*-stem combination forms a well defined set of adjectives as in (7) that is structurally distinct from the phonologically similar but morphologically simplex adjectives as in (8). Second, and more importantly, it needs to learn that the a-adjectives thus formed cannot be used attributively, which is the main problem at hand.

We will not review all the similarities between a-adjectives and locative participles 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 2011a, Goldberg 2011). More important, 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) is not at all attested in our child directed English corpus, which removes its utility in establishing the distributional properties of a-adjectives. Likewise, attested usage examples from the web, which previous researchers have used to study the distributional properties of the a-adjectives, can not be assumed to be available to the learner without an evaluation of the child directed language data.

However, one of the diagnostics has proven very robust and as we shall see in section 4, is abundantly attested in the input to English learning children. A class of adverbs such as *right*, *well*, *far*, *straight* and so on, which expresses the meaning of intensity or immediacy, can be used to modify a-adjectives. Following Bruening (2011a), we collectively refer to these structures as *right*-type modification.

(9)   a. I was well/wide awake at 4am
       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 that we surveyed although one can find attested examples on the web. But such adverbial modification can not appear at all with typical adjectives (10a–10b), while they are compatible with both locative particles (10c) and PPs (10d):

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

Despite these distributional similarities, we are somewhat reluctant to label a-adjectives as locative particles or prepositional phrases. Prepositional phrases are, of course, phrasal while a-adjectives are single words. Locative particles are single words but they are morphologically simplex while also appear to form a closed list. By contrast, a-adjectives are morphologically well structured and appear to be open-ended as illustrated in (6). 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 section 4, we show that the child learner can make use of the distributional evidence reviewed here to to establish the properties of the a-adjectives. For the moment, we turn to a reassessment of Boyd and Goldberg's (2011) statistical preemption proposal in detail.

3   On Statistical Preemption

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

3.1   Formal and Empirical Problems

The intuition for 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; see Pinker (1989) for a review. 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 $\bar{P}$ given the learning data thus presented. Boyd and Goldberg offer no suggestion 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 disassociated from psychological learning mechanisms.[1]

More importantly, and more empirically, statistical preemption makes a set of incorrect

predictions regarding the outcome of learning when tested on a large sample of child directed English. Our 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 concerning the usage patterns of adjectives, and will be supplemented by a 4.3 million word text corpus of child directed English. These corpora include 12 a-adjectives (see also section 4 below for the special case of *alive*):

(11)   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 so as to occur in a modest sample of child directed speech, none is frequent enough such that its absence of attributive usage would be conspicuous, which indeed has been suggested as another strategy for indirect negative evidence (e.g., Stefanowitsch 2008). In this 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 to be used in both constructions. But only one of the 12 a-adjectives (*afraid*, with a frequency of 73) falls in this higher frequency range, and many of the other 11 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: 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, we searched for all adjectives that appear predicatively in relative clauses (e.g., *do you have anything that is sharp*?). A total of 65 utterances are found, which involve 46 unique adjectives. Now if these adjectives appear exclusively in relative clauses, then their attributive usage must be preempted according to the statistical preemption hypothesis. We then searched for the attributive usages of the 46 adjectives in the parsed corpus but found only 28 of them. Below we give the remaining 18 adjectives used exclusively in relative clauses but not in noun phrases:

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

Those in italics (6 in all) are indeed adjectives that resist attributive use, including three a-adjectives, but none of the other 12 are, leading to a false positive rate of 66%. Meanwhile, of the 12 genuine a-adjectives in the 180,000 utterance corpus, 9 are not identified by statistical preemption, with a false negative rate of 75%.

It should be pointed out that 180,000 child directed utterances are not a sufficiently large sample of the input to the learner. We then turned to a 4.3 million word child directed text corpus from the CHILDES database (MacWhinney 2000), which roughly corresponds to a year's input data for some English children (see section 4 below) and includes the data from which the parsed corpus (Pearl & Sprouse 2013) is derived. If statistical preemption

works effectively, then the 12 adjectives falsely identified as attributively resistant may be found in prenominal positions in the much larger database. Unfortunately this is not the case: 8 adjectives (*annoying*, *careful*, *interested*, *jolly*, *torn*, *washable*, *wrapped*) still fail to appear prenominally in the larger database and would remain false positives. Increasing the size of the corpus will no doubt increase the number of adjectives that appear exclusively in relative clauses: these would be false positives under the statistical preemption hypothesis. Similarly, we searched for the 12 a-adjectives (11) in the 4.3 million word text corpus. We wanted to see how many of these would be used in a relative clause such that their attributive use can be preempted. Only one more (*asleep*) joins the ranks of the three identified in (12), and the remaining 8 are still unaccounted for.

Our corpus analysis also suggests that the probabilistic aspect of statistical preemption is unlikely to play any role in actual language learnng with respect to this case. As the search results show, adjectives are very rarely used in relative clauses: only 65 out of over 180,000 utterances (or 0.0003%), which are further spread over 46 unique adjectives. In other words, almost every adjective is used in a relative clause exactly once if at all. The child will not be able to accumulate evidence for $\bar{P}$ over multiple attestations so that the alternative hypothesis P is gradually weakened.

Finally, the statistical preemption hypothesis makes an additional prediction for **typical** adjectives which, as far as we 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 ($\bar{P}$), then the relative clause use of an adjective ($\bar{P}$) ought to be blocked if the learner only witnesses 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 will 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, which should be prevented from appearing 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. Indeed, statistical preemption would mislead the learner to falsely identify a small but non-negligible number of adjectives as attributively resistant (see 12), and 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 to reassess the role of statistical preemption. In all three experiments, adult English speakers consistently and strongly resist the attributive use for 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 witness various ways in which these adjectives are used so as to determine whether their attributive usage in a natural production task is affected as a result. Experiment 1 establishes a baseline performance. Without exposure to any usage examples, subjects are significantly less likely to use novel a-adjectives at-

tributively than novel non-a-adjectives (e.g., *chammy*). 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 in the 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 are in relative clauses while others are not. Nevertheless, all adjectives show considerably lower attributive usage rates than novel non-a-adjectives and in fact at a level similar to the existing a-adjectives. Boyd and Goldberg conclude that relative clause usage has preempted the attributive form while also allowing the subject to generalize over the entire class of novel a-adjectives. Experiment 3 contained the exposure patterns such as *the hamster that's ablim and proud of himself moved to the star*: subjects then appear to treat *ablim* more like a typical adjective in allowing for attributive use. Boyd and Goldberg suggest that the additional modification (*proud of himself*), which independently prohibits attributive use, discounted the preemptive effect from the relative clause.

There is an alternative account of these findings that Boyd and Goldberg overlooked, one which is also consistent with the distributional analysis of a-adjectives reviewed in section 2. That is, the subjects in Boyd and Goldberg's experiments are simply 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 remaining syllable after the segmentation of the *a-* schwa follows the phonotactics of English and could be interpreted as a potential (and unknown) word to describe some property of the animals used in the experiments.[2] 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, p67). 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*, *alert*, etc. 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 like words such as *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, that even without exposure to any usage pattern of the novel a-adjectives, subjects spontaneously limit their attributive use to a level significantly below the novel non-a-adjectives. Experiment 2 presents the subjects with the novel a-adjectives in relative clauses. This strengthens the morphologically based analysis (Experiment 1) that the subject 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 will encourage relative clause usage in the subject's production, 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 of a-adjectives, which can be extended to novel items; see (6). This,

however, begs the question as of how the speakers of English acquire these properties about the a-adjectives in the first place.

Regarding Experiment 3, our 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 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 the conjoined elements. For instance, in the 180,000 parsed child directed English sentences (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*) etc. 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 relative clauses is already extremely low (62 out of 180,000): only one involves conjoined adjectives (*those are special kinds of hairs that are sharp and pointy*). To expect an a-adjective in the relative clause in conjunction with another adjective followed by yet another modification that independently avoids attributive use, e.g., *ablim and proud of himself* as in Experiment 3, is next to impossible 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. We now show that the positive evidence for the grammatical properties of a-adjectives (section 2) is robustly attested in child directed English to guide language acquisition.

4 Generalization with positive evidence

Our 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; the resistance to attributive usage in the latter classes of linguistic units can be extended to the a-adjectives. However, not all a-adjectives are used in a PP like context in the child directed input; an independently motivated principle of generalization is necessary for the learner to extend the property to attributive resistance to the entire class of a-adjectives.

Our empirical study focuses on the distributional properties of a-adjectives in a 4.3 million child directed input corpus from the CHILDES database (MacWhinney 2000).[3] While more data would be desirable, this is not a trivial sample: it corresponds to more than a year of speech for children at the lowest socioeconomic level, more than a half for working class family children, and more than a third for professional family children (Hart & Risley 2003). To the extent that all speakers seem to learn the attributive prohibition on a-adjectives, our corpus provides a representative sample of the language acquisition data.

Boyd and Goldberg did not provide any analysis of child language, nor are we aware of any previous acquisition studies of a-adjectives. To address this issue, we searched

through an approximately 1.9 million word corpus for 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 a-adjective is found.[4] 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

We first extracted the attested a-adjectives from the child directed input. To identify the morphological structure of the 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 children's morphological knowledge across languages is generally highly reliable (Guasti 2004). English learner's word segmentation errors such as *there are three dults in our family* and *I was have* (Brown 1973, Peters 1983, Yang 2004) 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 the a-adjective stems are highly frequent and thus very likely to be part of a young child's vocabulary. For our quantitative analysis, we processed the corpus data with a part of speech tagger to extract all the adjectives. We then segmented off a word initial unstressed schwa *a* without violating the phonotactics of English; the results are presented below:

(13)   a.   Containing stems: afraid, <u>awake</u>, <u>aware</u>, ashamed, <u>ahead</u>, alone, <u>apart</u>, <u>around</u>, asleep, alike, <u>away</u>, <u>across</u>
        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 disjoint classes. All the items in (13a) contain a stem: all are a-adjectives. None of the items in (13b) contains a stem: none are a-adjectives and all can be used attributively. Thus, the morphological criterion provides a well defined membership for the a-adjectives. Second, we have searched for *right*-type modification usage for the non-a-adjectives in (13b).[5] Not a single instance such as those in (10a–10b) is 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 (13a) show robust usage with *right*-type modification in 8 out of the 12 types (shown underlined in 13a). We provide an example for each a-adjective below:

(14)   are you wide awake?
        I'm well aware of my shortcomings thank you
        go right ahead.
        it fell right apart on you.

> turn right around.
> finish the book right away.
> he fell fast asleep.
> we are coming right across.

The number of attested examples ranges from a handful (for *apart*) to over a hundred (for *away*). While our 4.3 million words corpus is not a trivial sample, positive evidence for the distributional properties of a-adjectives will be even more robust in larger samples of the primary linguistic data. We further note that the corpus contains numerous instances of *right*-type modification with locative particles (10c) as well as prepositional phrases (10d): *right here* and *right there* appear over 3,000 times, *right up/over/on* are in the hundreds, *right off/down/under* dozens, etc. 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 8 out of the 12 members of the a-adjectives in the input (13a) 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 41 a-adjectives in (7) that prohibit attributive use. Yet only 12 are attested in roughly a year's worth of child directed English, and 34 appear at all in a 50 million word spoken American English corpus (Brysbaert & New 2009). While that corpus only provides a word frequency list and does not make the text available, it is almost surely the case that only a fraction are paired with *right*-type modification, the signature evidence that relates a-adjectives to locative particles and prepositional phrases.

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 the learner does not witness enough positive instances, it 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 wait to witness the past tense form of every single verb.

4.2   Generalization

For concreteness, consider the Tolerance Principle (Yang 2005), a formal learning model of what counts as sufficient evidence to warrant generalization. Suppose a linguistic rule is applicable to N lexical items, out of which N_e are exceptions and do not follow the rule. The Tolerance Principle asserts that for the generalization to be productive (i.e., can tolerate the exceptions), N_e must not exceed $N/\ln N$. The 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), This allows for a cost/benefit calculus of processing complexity, with which the tolerable number of exceptions for a productive rule can be analyzed (see Yang 2005 for details).

We illustrate the application of the Tolerance Principle with two simple examples. First, consider the acquisition of English past tense. Suppose a typical English speaker knows $N_e = 120$ 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 $N/\ln N \geq 120$. 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 on novel items (Berko 1958).

Consider then an example from syntax. In a well-known critique, 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 precedes the noun:

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

Note that the problem of exceptions in (15) needs to be resolved under any theory of learning and the assumption of parameters is not essential. For instance, in a non-parametric 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 of the language places the adjective after it. To this end, we analyzed a corpus of approximately 200,000 child directed French sentences in the CHILDES database (MacWhinney 2000). The data is processed with a part of speech tagger to extract adjacent adjective and nominal pairs, which are then subject to manual inspection. All adjectives are lemmatized. In all, 20 exceptional adjectives appear in a prenominal position: if the postnominal order is to be productive, there needs to be approximately 80 typical adjectives to counterbalance the examples ($100/\ln 100 = 22 > 20$). Indeed, about half way into the corpus we already located at least 120 unique postnominal adjectives. A larger corpus can only increase this count, since the exceptional prenominal adjectives form 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. There are $N = 12$ attested a-adjectives, which are composed of the prefix *a-* and a stem. Out of these, the majority (8 out of 12) show explicit distributional evidence of patterning with locative particles. The learner has not encountered any explicit evidence concerning the remaining 4: the very worst case would be that they are exceptions to the general pattern detected so far (i.e., they can be used attributively). Even so, the generalization would still remain productive because 4 exceptions fall below the tolerated threshold of $5 = (12/\ln 12)$. 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 4 a-adjectives will inherit the property of attributive resistance from the 8 members with positive evidence, just as the child can extend the property of add "-d"

from verbs attested in past tense to verbs which have never appeared in past tense, once the rule of "add -d" has been established as productive as discussed earlier. Together with the well defined morphological criterion that a-adjectives consist of *a-* and a legitimate stem, the English speaker can acquire the properties of the a-adjectives and readily extend them to novel items as in (6) as well as in Boyd and Goldberg's 2011 experimental studies.

5 Summary

In this paper, we 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, we have used child directed English input to argue that a-adjectives can be reliably identified with locative particles and locative phrases, under an independently motivated principle of generalization (Yang 2005). The strategy of statistical preemption (Boyd & Goldberg 2011) is neither sufficient nor necessary.

We 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 which 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 the problem of child language acquisition. If so, the learner's task would presumably be further simplified. Indeed, our proposal deals with a worst case situation, that the child is to acquire the properties of the a-adjectives with the most basic and traditional use of distributional learning across contexts (Harris 1951): 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 (8 out of 12) will suffice to derive the same learning outcome.

We 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 for data. For instance, previous researchers (Boyd & Goldberg 2011, Bruening 2011a, 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 sort of expressions found in large linguistic 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 is found in actual child directed English.

It is widely known that only a small fraction of the linguistically possible forms will be 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. Almost by necessity, the properties of the a-adjectives as a class can only be acquired on the basis of very few 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, and the learner must be well equipped to connect seemingly disparate phenomena so they can learn them together, by generalizing aggressively and early on.

Notes

[1]For a recent empirical study, see Perfors et al. (2010) and Villavicencio et al. (2013).

[2]We thank an anonymous reviewer for raising this possibility.

[3]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 CHILDS project recommends a consistent coding format for children's speech, with each utterance marked as "*CHI", not all transcripts follow this convention so it is impossible to automatically distinguish child from child directed speech from these files. Thus, we only included transcription files which marked children's utterances with "*CHI:" and extracted both child and child directed speech from these files, which nevertheless retain the vast majority of the American English acquisition data in the public domain (2133 out of the 2279 transcripts files).

[4]We did find one instance of *alive* used attributively (*alive things walk*), but we also found similar patterns by adults (e.g., *this is the alive one*, *A Thumping! A Bumping! A Wild Alive Scratching!* from the well known Dr. Seuss book). Presumably this adjective, at least for some speakers, 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. We exclude this word from further consideration.

[5]We thank an anonymous reviewer for this suggestion.

References

BAKER, CARL L. 1979. Syntactic theory and the projection problem. *Linguistic Inquiry* 10.533–581.

BERKO, JEAN. 1958. The child's learning of English morphology. *Word* 14.150–177.

BERWICK, ROBERT. 1985. *The acquisition of syntactic knowledge*. The MIT Press, Cambridge, MA.

BOCK, J KATHRYN. 1986. Syntactic persistence in language production. *Cognitive psychology* 18.355–387.

BOLINGER, DWIGHT LE MERTON. 1971. *The phrasal verb in english*. Cambridge, MA: Harvard University Press.

BOULDIN, JOHN MICHAEL. 1990. *The syntax and semantics of postnominal adjectives in english*. 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: Basil 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.

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, 153–186. 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 again: Response to Goldberg. Available at http://lingcomm.blogspot.com/.

BRUENING, BENJAMIN. 2011b. A-adjectives are PPs, not adjectives. Available at http://lingcomm.blogspot.com/. Online: `http://lingcomm.blogspot.com/2011/03/adjectives-are-pps-not-adjectives.html`.

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* 41.977–990.

CAMPBELL, AIMEE L, and MICHAEL TOMASELLO. 2001. The acquisition of English dative constructions. *Applied Psycholinguistics* 22.253–267.

CHATER, NICK, and PAUL VITÁNYI. 2007. Ideal learning of natural language: Positive results about learning from positive evidence. *Journal of Mathematical psychology* 51.135–163.

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: Erlbaum.

CONWELL, ERIN, and KATHERINE DEMUTH. 2007. Early syntactic productivity: Evidence from dative shift. *Cognition* 103.163–179.

COPPOCK, ELIZABETH. 2008. *The logical and empirical foundations of Baker's Paradox*. Stanford University dissertation.

DAVIES, MARK. 2008. Corpus of Contemporary American English (COCA): 410+ million words, 1990-present. available at http://www.americancorpus.org/. 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 36th annual boston university conference on language development*, ed. by A. Biller, E. Chung, and A. Kimball, vol. 1, 164–176. Somerville, MA: Cascadilla Press.

FODOR, JANET DEAN. 1989. Learning the periphery. *Learnability and linguistic theory*, ed. by R. J. Matthews and W. Demopoulos, 129–154. 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. Lawrence Erlbaum Associates.

FODOR, JANET DEAN, and WILLIAM GREGORY SAKAS. 2005. The subset principle in syntax: Costs of compliance. *Journal of Linguistics* 41.513–569.

GERKEN, LOUANN, and BONNIE J MCINTOSH. 1993. Interplay of function morphemes and prosody in early language. *Developmental Psychology* 29.448–457.

GOLDBERG, ADELE E. 2011. Are a-adjectives like afraid prepositional phrases underlying and does it matter from a learnability perspective? Unpublished manuscript. Princeton University.

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

HARRIS, ZELLIG S. 1951. *Methods in structural linguistics*. Chicago: University of Chicago Press.

HART, BETTY, and TODD R RISLEY. 2003. The early catastrophe: The 30 million word gap by age 3. *American Educator* 27.4–9.

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* 47.637–669.

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

LARSON, RICHARD K, and FRANC MARUŠIČ. 2004. On indefinite pronoun structures with aps: Reply to Kishimoto. *Linguistic Inquiry* 35.268–287.

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

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

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

MARCUS, GARY; 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.

MARCUS, GARY F. 1993. Negative evidence in language acquisition. *Cognition* 46.53–85.

MAZURKEWICH, IRENE, and LYDIA WHITE. 1984. The acquisition of the dative alternation: Unlearning overgeneralizations. *Cognition* 16.261–283.

NEWMEYER, FREDERICK J. 2004. Against a parameter-setting approach to typological variation. *Linguistic Variation Yearbook* 4.181–234.

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

PENKE, MARTINA, and MARION KRAUSE. 2002. German noun plurals: A challenge to the dual-mechanism approach. *Brain and Language* 81.301–311.

PERFORS, AMY; JOSHUA B. TENENBAUM; and ELIZABETH WONNACOTT. 2010. Variability, negative evidence, and the acquisition of verb argument constructions. *Journal of Child Language* 37.607–642.

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* 39.633–651.

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, 111–130. Stanford: CSLI.

RANDALL, JANET. 1990. Catapults and pendulums: the mechanics of language acquisition. *Linguistics* 28.1381–1406.

RAUH, GISA. 1993. On the grammar of lexical and non-lexical prepositions in english. *The semantics of prepositions*, ed. by Michel Aurnague and Laure Vieu, 99–150. Walter de Gruyter.

SAG, IVAN A; GERALD GAZDAR; THOMAS WASOW; and STEVEN WEISLER. 1985. Coordination and how to distinguish categories. *Natural Language & Linguistic Theory* 3.117–171.

SALKOFF, MORRIS. 1983. Bees are swarming in the garden: a systematic synchronic study of productivity. *Language* 59.288–346.

STEFANOWITSCH, ANATOL. 2008. Negative entrenchment: A usage-based approach to negative evidence. *Cognitive Linguistics* 19.513–531.

STVAN, LAUREL SMITH. 1998. *The semantics and pragmatics of bare singular noun phrases*. Northwestern University dissertation.

SWINNEY, DAVID, and ANNE CUTLER. 1979. The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior* 18.523–534.

TENENBAUM, JOSHUA B., and THOMAS L. GRIFFITHS. 2001. Generalization, similarity and bayesian inference. *Behavioral and Brain Sciences* 24.629–640.

TOMASELLO, MICHAEL. 2000. Do young children have adult syntactic competence? *Cognition* 74.209–253.

VALIAN, VIRGINIA; STEPHANIE SOLT; and JOHN STEWART. 2009. Abstract categories or limited-scope formulae? the case of children's determiners. *Journal of Child Language* 36.743–778.

VILLAVICENCIO, ALINE; MARCO IDIART; ROBERT C BERWICK; and IGOR MALIOUTOV. 2013. Language acquisiiton and probabilistic models: Keeping it simple. *Proceedings of the 51st annual meeting of the association for computational linguistics*, 1321–1330.

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* 8.451–456.

YANG, CHARLES. 2005. On productivity. *Linguistic Variation Yearbook* 5.333–370.

YANG, CHARLES. 2013. Ontogeny and phylogeny of language. *Proceedings of the National Academy of Sciences* 110.6324–6327

ZIPF, GEORGE KINGSLEY. 1949. *Human behavior and the principle of least effort: An introduction to human ecology*. Cambridge: Addison-Wesley.