

Constructional Fast Mapping

DEVIN CASENHISER and ADELE E. GOLDBERG
Princeton University

0. Introduction

In learning a language, children must generalize over the utterances they hear so that they can creatively produce and understand utterances they have never heard before (Chomsky, 1957). Statistical properties of language have been demonstrated to be key in enabling the necessary formal generalizations (Childers and Tomasello, 2002; Gomez, 2002; Saffran, 2001; Saffran et al., 1996). However, no one has investigated experimentally how children come to learn mappings between novel phrasal forms and novel meanings: exactly the task that presents itself to learners in natural settings. The results reported here demonstrate that with surprisingly minimal input, children are able to generalize beyond their experience; moreover, the high token frequency of a single exemplar, which has been found to exist in naturalistic input to children, is demonstrated to have a facilitatory effect. While the fast mapping might be taken as an indication of innate, language-specific knowledge, we argue on the contrary, that parallel results reported in a non-linguistic categorization task indicate that an appeal to an innate language faculty is not required in this particular domain.

Children learning language must come to know correlations between phrasal patterns and meanings, so that when they hear novel verbs in utterances such as *She text-messaged him the directions*, they are able to discern in a general way what that new utterance means: in this case something like “*She GAVE him the directions using text messaging*” (Goldberg, 1995; Landau and Gleitman, 1985). This is something on which all linguistic and psycholinguistic theories agree; there exist correlations between phrasal forms and meanings. The results reported here demonstrate that with quite minimal input, children can learn these correlations and can generalize on the basis of them. Moreover, we demonstrate that the learning is also critically facilitated by a high number of instances of a single verb type. This sort of statistical “skewing” of the input, where a restricted subset of types of utterances accounts for the preponderance of total utterances, is exactly what is found in naturalistic speech to children for a number of types of language patterns (Cameron-Faulkner et al., to appear; Diessel, 2002; Thompson

and Hopper, 2001). Thus the learning mechanism demonstrated here may have a fairly general utility, allowing learners to get an initial fix on many types of form-meaning mappings in language; this potentially offers a way that learners can crack into the system of learning to use language in an infinitely creative way.

Previous work has focused almost entirely on the question of whether mappings between phrasal patterns and meanings have been acquired at a certain age rather than on particular statistical factors that facilitate or inhibit the learning of the mappings. J. Childers and M. Tomasello (2001) is the only training study that has found a facilitating factor, namely the use of pronouns instead of full NP arguments in the acquisition of the English transitive construction; K. Abbott-Smith, E. Lieven, M. Tomasello, *Developmental Science* (in press) attempted to look for other factors but found null results. (Fisher, 1996; Naigles, 1990; Tomasello, 2000). To some extent the lack of experimental work is due to the fact that many researchers believe that these kinds of mappings must be innate, therefore essentially eliminating the need to learn them from the input; they have been thought to be hard-wired into a biological language mechanism called universal grammar (Nowak et al., 2002). If, however, the statistical nature of the input can be demonstrated to be critical to the domain of meaning as well as form, then it would reduce the necessity of invoking innate universal grammar to account for the mappings. This would be advantageous insofar as accounts that rely on a universal grammar fail to predict the item-based, or bottom-up way that children learn language (Tomasello, 2000, 2003). Moreover, recognizing that form-meaning mappings can be learned from the input allows for substantial variation in the mappings across languages, variation that in fact has been argued to exist (Bowerman and Brown, to appear; Bowerman, 1990). At the same time, it would raise the further question of whether the learning mechanism involved is specific to language, or whether the learning strategy is instead general to cognition.

An analysis of naturalistic data reveals that a single verb typically accounts for the lion's share of tokens of each of several simple patterns in the input speech of mothers to young children (Goldberg 1998; Goldberg et al., to appear; Ninio, 1999). For example *put* fills the verb slot in roughly 40% of the instances of the phrasal pattern, <Subject - Verb - Object - Locative Phrase>; *give* fills the verb slot in roughly 20% of the phrasal pattern <Subject - Verb - Object1 - Object2>; and *go* fills the verb slot in roughly 40% of the phrasal pattern <Subject - Verb - Locative Phrase>. These frequencies are strongly skewed in that the constructions have been found to occur with 43, 13 and 39 different verb types in the same corpus sample, respectively. Moreover, the meaning associated with highly frequent verbs like *put*, *give* and *go* has been independently claimed to be the semantic prototype of the meaning of the phrasal pattern in which these verbs occur so frequently (Goldberg, 1995; Pinker, 1989). For example, the verb *put* and the phrasal pattern with which it is associated convey a 'caused motion' meaning; *put* means 'to cause something to move to a location'. Likewise, *put*'s phrasal pattern, <Subject - Verb - Object - Locative Phrase>, also conveys a

caused motion meaning. That the phrasal pattern does in fact suggest such a meaning becomes apparent by considering a sentence such as *Pat mooped the feather onto the table*. Although it is not certain exactly what *mooped* means, it seems clear from the sentence that the feather in question has somehow found its way to the table with the help of Pat.

Since the input is typically structured such that a small subset of types account for the preponderance of utterances, we hypothesized that the existence of a high frequency exemplar facilitates association of a meaning with a phrasal pattern. To test this, we created a novel, non-English phrasal pattern and paired sentences instantiating this pattern with a film of various scenes in which a puppet or toy object appeared on the scene in some way. The intended meaning of the phrasal pattern was accordingly one of appearance. We then divided subjects into three groups: 1) the high frequency group watched the film and heard a corresponding set of sentences in which one novel verb occurred in half of the sentences, 2) the balanced group watched the same film and heard a set of sentences in which each of the novel verbs occurred with roughly equal frequency, and 3) the control group watched the same film with the sound turned off. Subjects were tested with a forced-choice comprehension task to determine if they were able to extend the meaning of the new phrasal pattern to which they were exposed during the experiment to correctly choose new scenes of appearance paired with new novel verbs over foil scenes we created that were not scenes of appearance but were similar to the scenes of appearance in every other way.

1. Methods

1.1. Subjects

1.1.1. Experiment One

51 native English speaking children aged 5-7 (mean = 6;4) were recruited from two elementary schools in Champaign-Urbana.

1.1.2. Procedure

A single training film was prepared that contained eight clips of puppets performing various actions. The same film was presented twice to each subject (a total of 16 video clips). In the non-control conditions, we paired each clip in the film with audio descriptions of the scene and arranged the words in the description according to a novel phrasal pattern that we created. The novel pattern involved two known nouns along with a nonsense verb and were arranged in the form <noun phrase₁ - noun phrase₂ - nonsense verb + o >. For example, given a video clip in which a spot appeared on the king's nose, the corresponding sentence was 'The spot the king moopo-ed'. At the beginning of each scene, subjects heard a present tense version of the sentence, and heard a corresponding past tense version of the sentence at the end of the scene.

The meaning of the phrasal pattern was that of appearance, (a meaning novel for English phrasal patterns): the entity named by the first noun phrase comes to exist in the place named by the second noun phrase, according to the action

encoded by the verb. For example, the intended meaning for the sentence *the sailor the pond neebod* was ‘the sailor sailed onto the pond from out of sight’ as opposed to, for example, ‘the sailor sailed (around) the pond’.

Subjects were randomly assigned to one of three conditions that varied the frequencies of input exemplar nonsense verbs; the number of different nonsense verbs (5) and the overall number of examples (16) were held constant. Training in the balanced condition consisted in hearing expressions with five nonsense verbs paired with video clips, each occurring 1 or 2 times (1-1-2-2-2); the high frequency condition also heard 5 nonsense verbs used in the novel phrasal pattern, but one nonsense verb was heard with high frequency (1-1-1-1-4). Subjects in the control condition saw the identical film of 8 clips shown twice (for a total of 16 clips) but heard no language. The training film was played twice for each of the three groups. The total length of the training session was less than three minutes.

The test was a forced choice comprehension task: subjects saw two *new* film clips presented side-by-side on the screen and heard a sentence describing one of the clips (sample film clips are available at <http://www.linguistics.uiuc.edu/casenhis/fastmap/>). Sentences included 7 test trials with the novel phrasal pattern and *new* novel verbs; interspersed were 5 filler trials with other *new* novel verbs in the familiar transitive pattern. Each test film clip pair showed the same entity involved in a similar action, but only in one did the entity appear on the scene within the clip (e.g., in one case, a sailor sails in on a boat from off the screen; in the paired foil clip the sailor sails around in a boat on screen). Subjects were then asked to point to the film clip that corresponded to the description that they heard. Responses were coded for accuracy. Any difference among groups can only be attributed to a difference in the linguistic input that subjects were exposed to, as all three conditions watched exactly the same video.

1.1.3. Results

The results of the experiment show that after only three minutes of training, children in the balanced and high frequency groups learned to associate a novel meaning with a novel phrasal pattern. Moreover, the high frequency group performed significantly better than the balanced group, thus confirming our hypothesis that learning is particularly facilitated when one verbal token accounts for the majority of utterances as seen in figure 1.

Constructional Fast Mapping

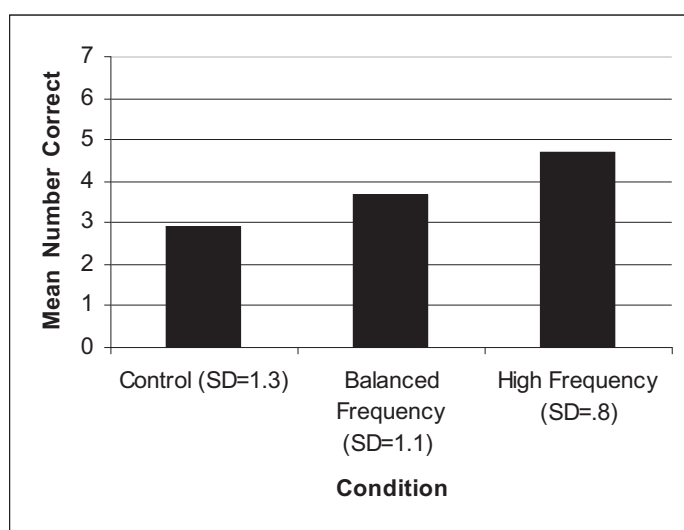


Figure 1 Mean number of times subjects were able to correctly match the scene of appearance to the novel phrasal pattern.

An ANOVA confirmed a significant main effect for group, $F_{2, 48} = 11.57$, $P < .001$. Planned comparisons analysed with Fisher's PLSD show that both the high frequency and the balanced groups performed significantly better than the control group ($P < .001$ and $P < .05$ respectively). Moreover, the high frequency group performed significantly better than the balanced group ($P < .01$).

Similar results have been found for adults (Goldberg, Casenhiser, and Sethuraman 2004); in the adult experiment, each nonsense verb was matched with a specific type of action, so that the two training conditions saw slightly different films. In addition, the control condition in the previous experiment did not watch the film but went straight to test.

It is possible that the quick learning of the mapping could be taken as an indication that the particular mapping is a part of universal grammar and is innately available. A mapping between subject and thing coming to exist, on the one hand, and displaced noun phrase and location on the other, could be added to the set of mapping principles sometimes claimed to be universal. However, we know of no language that has a general mapping that encodes “appearance” in this way. Thus, given its cross-linguistic rarity, there is no independent reason to believe that the particular generalization learned in the study reported here is innately available.

Moreover, there is reason to suspect that the learning *mechanism* is not specific to language, but is general to cognition insofar as work in the non-linguistic category literature has found a parallel facilitatory factor to that reported here. That is, there is an advantage to training on low-variance input, and on prototypical instances before more varied input in the learning of non-linguistic categories (Elio and Anderson, 1984; Homa et al., 1991; Nosofsky, 1988; Rosch and Mervis, 1975). For example, there is a strong correlation between the frequency with which a token occurs and the likelihood that it will be considered a

prototype by the learner (Nosofsky, 1988; Posner and Keele, 1968; Rosch and Mervis, 1975). Homa, Dunbar and Nohre (1997) found that token frequency was an important variable at early and intermediate stages of category learning, with increased token frequency facilitating category learning. In learning generalizations about dot patterns, Posner, Goldsmith and Welton (1967) demonstrated that the rate at which subjects classified patterns correctly was a direct function of the amount of distortion from their respective prototypes: the less variability or distortion, the faster the category was learned.

Elio and Anderson's (1984) non-linguistic category learning experiment set up two conditions: in the "centered" condition, subjects were initially trained on more frequently represented, more prototypical instances, with the study sample growing gradually to include the full range of members in the category; in the "representative" condition, subjects were trained on a fully representative sampling from the start. Categories were learned more accurately in the centered condition, yielding better typicality ratings and accuracy during the test phase on new instances.

We performed a second experiment with a parallel design to test the relationship to non-linguistic categorization. We created a random dot pattern (with 10 dots) to be used as a prototype as well as 4 systematic variations from the prototype pattern. Subjects in the high frequency group saw twice as many instances of the prototype dot pattern as any of the other dot patterns. Subjects in the balanced group were not given this preferential training with the prototype; instead, they saw a more balanced distribution of the prototype pattern in comparison to the other dot patterns. Subjects were again tested with a forced choice to determine if they were able to distinguish a *new* variation of the prototype from a dot pattern generated randomly. New variations used at test differed from the prototype to the same degree as the variations used in training.

1.2. Experiment Two

1.2.1. Subjects

28 University of Illinois undergraduate students.

1.2.2. Procedure

A prototype dot pattern consisting of 10 uniformly-sized white stars scattered across a 30 by 40 grid was prepared. From this prototype, a set of 10 close variations of the prototype were created by moving 4 of the stars 1-3 grid squares in a random direction. The number of squares each star moved was also chosen randomly. 6 foil patterns were created by moving 4 of the stars 3-6 grid squares in a random direction. Again the number of squares and the stars that were moved were chosen randomly. The training set included the prototype slide plus four of the close variation star patterns chosen randomly (for the sake of clarity, these will be referred to as variations 1-4). The test set included each of the 6 remaining close variation star patterns (variations 5-10) placed beside of the 6 foil star patterns.

Constructional Fast Mapping

Slides were ordered randomly and each was created with a different background color. Subjects viewed each slide on a computer screen for 3 seconds. To prevent the illusion of movement when one star pattern was presented immediately after another, each star pattern slide was separated from the following one by a black slide with a regular grid of moons.

Subjects were randomly assigned to one of two groups. The high frequency group saw a set of slides in which half (8) of the slides were the prototype slide. The remaining 8 slides were prototype variations 1-4 and were viewed twice each. The balanced group saw the prototype slide plus variations 1 and 2 shown four times each. Variations 3 and 4 were shown twice each. Both groups watched the slide show containing a total of 16 star pattern slides twice. The first star pattern in each group was the prototype, and the second star pattern in each group was one of the close variations (1-4). All other slides were randomly ordered. Each group's training lasted 2 minutes and 7 seconds.

The test was a forced choice task in which subjects were shown a slide containing two star patterns placed side-by-side. One of the star patterns was from variations 6-10 and the other was one of the foil star patterns. Subjects were asked to choose the pattern that was the same as the star pattern from the training portion of the experiment.

1.2.3. Results

The results demonstrate that subjects in the high frequency group were more accurate at test than those in the balanced frequency group, thus confirming the suggestion that learning of categories generally is facilitated when a prototype is encountered with high frequency as opposed to experience with the same variety of instances including the prototype, when the prototype does not account for the balance of items as seen in figure 2:

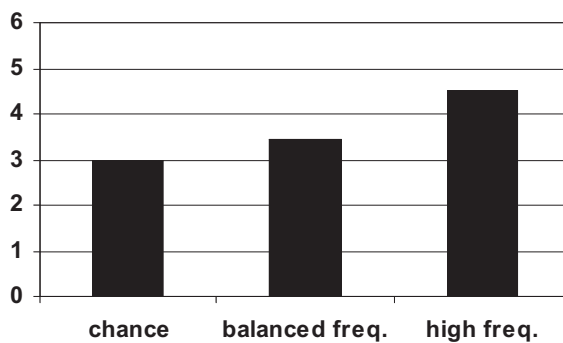


Figure 2: Mean number of times subjects were able to correctly match the new variation of the random dot pattern.

An ANOVA confirmed a significant main effect for group, such that the high frequency training condition performed significantly better than the balanced frequency training condition $F(1,27)=6.78, p < .02$. Chi-square test with 6 degrees

of freedom was performed to compare subjects' performance to chance. Subjects in the balanced group did not perform significantly above chance $p = .13$. Subjects in the high frequency group did perform significantly above chance $p < .01$.

2. Conclusion

In the case of language, we have found that pairings of novel phrasal patterns and novel meanings are generalized with remarkable speed. When they are instantiated predominantly by a single verb, they are generalized even more effectively. We have also reported evidence that the latter effect is paralleled in a non-linguistic categorization task, providing a strong indication that the learning mechanism is not specific to language. Since natural linguistic input tends to be skewed in this way, it seems that the associations of form and meaning that exist in languages do not need to be hard wired or universal—children are quite expert at learning the mappings, given general categorization strategies.

References

- Abeillé, Anne, and Danièle Godard. 1997. French word order and the word phrase distinction. Stanford University.
- Bowerman, Melissa. 1990. Mapping thematic roles onto syntactic functions: Are children helped by innate linking rules? *Linguistics* 28: 1253-1289.
- Bowerman, M., and P. Brown (eds.) To appear. *Crosslinguistic Perspectives on Argument Structure: Implications for Learnability*. Lawrence Erlbaum.
- Cameron-Faulkner, T., E. Lieven, and M. Tomasello. To appear. *Cognitive Science*.
- Childers, J., and M. Tomasello. 2002. Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental Psychology* 38: 967-978.
- Chomsky, Noam. 1957. *Syntactic Structures*. Berlin: Mouton.
- Diessel, Holger. 2002. The development of complex sentence constructions in English – A usage-based approach. University of Leipzig, Habilitation thesis.
- Elio, Renee, and John R. Anderson. 1984. The effects of information order and learning mode on schema abstraction. *Memory and Cognition* 12: 20-30.
- Fisher, Cynthia. 1996. Structural limits on verb mapping: The role of analogy in children's interpretations of sentences. *Cognitive Psychology* 31: 41-81.
- Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Goldberg, Adele E. 1998. The emergence of the semantics of argument structure constructions. In Brian MacWhinney (ed.) *The Emergence of Language*, 197-212. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Goldberg, Adele E., Devin M. Casenhiser, and Nitya Sethuraman. 2004. Learning argument structure generalizations. *Cognitive Linguistics* 14: 289-316.

Constructional Fast Mapping

- Gomez, Rebecca. 2002. Variability and detection of invariant structure. *Psychological Science* 13: 431-436.
- Homa, Donald, Sherry Dunbar, and Liva Nohre. 1991. Instance frequency, categorization, and the modulating effect of experience. *Journal of Experimental Psychology: Learning, Memory, & Cognition* 17: 444-458.
- Landau, Barbara, and Lila R. Gleitman. 1985. *Language and Experience: Evidence from a Blind Child*. Cambridge, MA: Harvard University Press.
- Naigles, Letitia. 1990. Children use syntax to learn verb meanings. *Journal of Child Language* 17: 357-374.
- Ninio, Anat. 1999. Pathbreaking verbs in syntactic development and the question of prototypical transitivity. *Journal of Child Language* 26.
- Nosofsky, Robert. 1988. Similarity, frequency and category representations. *Journal of Experimental Psychology: Learning, Memory and Cognition* 14: 54-65.
- Nowak, Martin A., Natalia L. Komarova, and Partha Niyogi. 2002. Computational and evolutionary aspects of language. *Nature* 417: 611-617.
- Pinker, Steven. 1989. *Learnability and Cognition: The Acquisition of Argument Structure*. Cambridge, MA: MIT Press/Bradford Books.
- Posner, Michael I., Ralph Goldsmith, and Kenneth E. Welton, Jr. 1967. Perceived distance and the classification of distorted patterns. *Journal of Experimental Psychology: General* 73: 28-38.
- Posner, Michael I., and Steven W. Keele. 1968. On the genesis of abstract ideas. *Journal of Experimental Psychology* 77: 353-363.
- Rosch, Eleanor, and Carolyn B. Mervis. 1975. Family Resemblances: Studies in the internal structure of categories. *Cognitive Psychology* 7.
- Saffran, Jenny R. 2001. The use of predictive dependencies in language learning. *Journal of Memory and Language* 44: 493-515.
- Saffran, Jenny R., Richard N. Aslin, and Elissa L. Newport. 1996. Statistical learning by 8-month-old infants. *Science* 274: 1926-1928.
- Thompson, Sandra A., and Paul J. Hopper. 2001. Transitivity, clause structure and argument structure: Evidence from conversation. In J. Bybee and P.J. Hopper (eds.) *Frequency and the Emergence of Linguistic Structure*, 27-60. Amsterdam: John Benjamins.
- Tomasello, Michael. 2000. Do young children have adult syntactic competence? *Cognition* 74: 209-253.
- Tomasello, Michael. 2003. *Constructing a Language*. Boston: Harvard University Press.

Devin Casenhiser

Adele Goldberg
Psychology Department
Green Hall
Princeton, NJ 08544

adele@princeton.edu