Rapid learning of word meanings from distributional and morpho-syntactic cues

Margherita De Luca (margherita.deluca@uniroma1.it)
University of Rome "La Sapienza", Department of Document studies, Linguistics, Philology and Geography, Piazzale Aldo Moro, 5
Rome, 00185 Italy

Gary Lupyan (lupyan@wisc.edu)
University of Wisconsin-Madison, Department of Psychology, 1202 W. Johnson Street
Madison, WI 53706-1611 USA

Abstract

What does it take to learn a new word? Many of the words we learn, we have learned from language itself – by encountering them in various informative contexts. Here, we investigate the limits of learning from context by studying how people learn new words from very sparse contexts, at the extreme, a context in which all content words are replaced by nonsense words. We find that participants exposed to even such extremely *sparse* contexts nevertheless learn something about the meaning of words embedded in those contexts. Performance tended to be better when knowledge was assessed by first directing people's attention to the part of speech of the target words.

Keywords: language; word learning; distributional semantics; syntactic bootstrapping.

Introduction

How do we learn the meanings of words? One way is by associating words with external referents. This is of particular importance in early word-learning when children's vocabulary is dominated by concrete nouns ("cup", "nose"), action verbs ("jump", "bark") and words used by children as imperatives ("more", "up"). As we become more mature language users, our vocabulary expands to contain a large proportion of items that do not have concrete referents (e.g., "terrible", "hope" and "fun"). These words must be learned through language itself.[1]

One way to learn word meanings through language is through explicitly provided definitions. For example, on encountering the sentence "The Celtics Coach was livid over the call, hurling an expletive at the officials.", one might look up "livid" to learn that a common meaning is "furiously angry". But we can also learn something about the possible meaning of livid through its context. Indeed, school-age children learn, on average, between 600 and over 3,000 new words per year (Nagy & Herman, 1984; Nagy, Herman, & Anderson, 1985). That these words are learned via direct instruction or through dictionaries is highly unlikely (Nagy & Anderson, 1984).

In this work we investigate the limits of learning from linguistic context. We already know that school-age children and adults are adept at learning from informative contexts such as the "livid" example above. We were curious whether learning something about a word's meaning is possible even when contexts are highly impoverished. At the extreme, we investigate learning meanings of nonce words from contexts in which all content words are replaced with nonce terms. How can people learn word meanings from contexts that themselves contain only nonce words? One way is through the use of morpho-syntactic cues.

The idea that people can learn words through morphosyntax is generally known as "syntactic bootstrapping". Preschool children can use both morphology and syntax to infer the meanings of novel words (Gleitman, 1990; A. E. Goldberg, 2003; Naigles & Swensen, 2007). In a classic demonstration, Brown (1957) showed 3-5 year old children images such as a pair of hands emptying an odd container of a novel slushy material. The children were then told that "in this picture you can see sibbing/a sib/some sib", and were asked to point to another instance of sibbing/a sib/some sib. The finding that children were more likely to point to a substance when presented with "some sib" and an object when presented with "a sib" suggests that they are learning (coarse) aspects of meaning from morpho-syntactic cues. In a more recent study, Naigles presented 2-year old children with the novel word "gorping" and two videos of novel actions, one causative and one non-causative. The sentence frame in which the verb was presented influenced which action children chose for the verb (Naigles, 1990). Although these experiments show that even young children can make use of morpho-syntactic cues to learn something about what the word means, they do not tell us about the limits of our ability to learn word meanings in this way. In these classic studies, the to-be-learned are marked, e.g., by being the only unfamiliar word in the utterance. The contexts that are used tend to be quite informative and the meanings being learned are constrained by the accompanying pictures or videos. Lastly, the assessment of what is being learned in such studies tend to be quite limited. The children in Brown's study could answer "correctly" simply by inferring that "sibbing" is an action, without learning anything more specific about its meaning.

But we know that people can do more. For example, even in a language with relatively simple morphology – English – a sentence like "The gostak distims the doshes" is surprisingly meaningful. We can infer that a "gostak" is doing something ("distimming") and it is doing it to the "doshes" (Ingraham, 1903). We may further infer that a gostak may be an animate

[1] Indeed, in the largest set of concreteness/abstractness norms (Brysbaert, Warriner, & Kuperman, 2014), words were defined as "abstract" if their meanings could not be communicated via direct reference or through action, but rather would have to be explained through language.

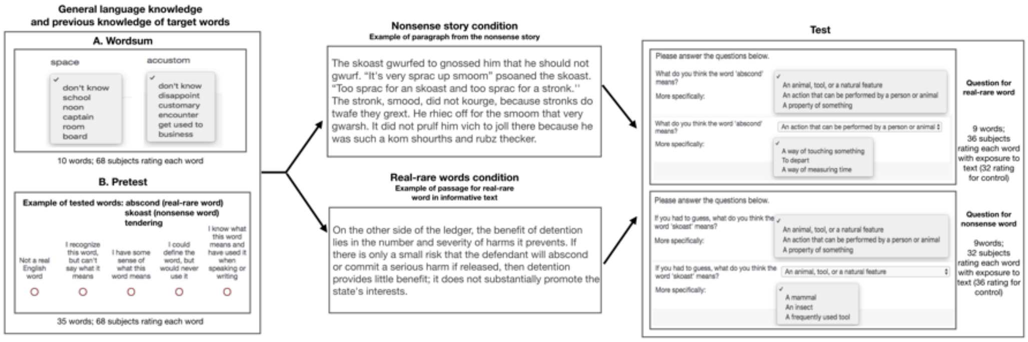

Figure 1: Schematic for Experiment 1.

agent and "distimming" is an action that the gostak is capable of performing. Remarkably, even without *any* referential context, exposure to sequences of such sentences appears to be sufficient for people to navigate an entire virtual world. Players of *The Gostak* (Muckenhoupt, 2001) quickly learn to *deave in the tavid dorl and gomb the stike*.

Another hint that language contains rich cues to word meanings comes from attempts to derive semantic representations from the ground up, by tracking word co-occurrences to learn that words occurring in similar contexts have similar meanings (Landauer & Dumais, 1997; Lany & Saffran, 2010; Levy & Goldberg, 2014; Mikolov, Chen, Corrado, & Dean, 2013; Redington, Chater, & Finch, 1998) – the so-called distributional hypothesis of word meaning (Firth, 1957). McDonald and Ramscar (2001) tested the distributional hypothesis by manipulating the linguistic context surrounding very rare or nonce words showing that judgments involving the target words are affected by the distributional properties of the manipulated linguistic environments.

Our main goal is to investigate the limits of people's ability to learn word meanings from linguistic context. We do this by exposing adult English speakers to contexts varying in informativeness ranging from fully informative contexts – passages of real English text containing real words unknown to most of our subjects, to highly sparse contexts that contain English morpho-syntactic cues (e.g., verb endings), but in which all content words have been replaced by nonsense words. We test people's knowledge of both part of speech information (the sole focus of much of the classic work on syntactic bootstrapping), *and* knowledge of more specific aspects of meaning.

Experiment 1

In experiment 1, we tested the role of the linguistic context in inferring word meanings for real, but rare English words (e.g., "kine") that were presented in informative contexts and for nonce words (e.g., "stronk") placed in highly *sparse* contexts which were stripped of almost all meaningful words.

Participants We recruited 114 participants from Amazon Mechanical Turk (52 Males of average age = 37; 62 Females of average age = 39). 68 of these participated in the main word-learning experiment (32 in the nonsense-story condition and 36 in the real-passages condition) and 46 participated in the salience-norming task.

Materials Participant were randomly assigned to one of two conditions: a nonsense-story condition and a real-passages condition. For the nonsense-story condition, all participants were exposed to the children's story *Why the cricket chirps*[2]. Participants were not provided with the story's title or any information about its content. Of the 604 total tokens in the story, 169 were open-class words. We replaced all the content words with nonce words taken from the ARC Nonword Database (Rastle, Harrington, & Coltheart, 2002). These words were created from orthotactically legal bigrams, onsets, and bodies. Fig. 1 shows a part of the resulting story. Of the 136 word types in the story that were replaced with nonce words, we selected 9 to serve as targets for later testing. Participants did not know ahead of time which words would be tested. The target words varied in frequency, occurring between 2 and 18 times, and parts of speech: 4 nouns, 3 verbs, 2 adjectives. Derivational and inflectional morphemes in the story were limited to a small number of cases (see Table 1 for morphological variation present for each target word).

For the real-passages condition, we matched each of the 9 target words in the nonsense-story condition with real, but rare English words that were unlikely to be familiar to our participants (e.g. "ratoon", "pronk, "rawky"; henceforth *real-rare* words). For each word, we selected 3-4 sentences in which the word appeared from the Corpus of Contemporary American English (COCA) and other online sources to serve as the context (see Fig. 1). Participants' word knowledge in all conditions was tested using "drill-down questions" designed to be sensitive to partial word knowledge. The first question provided three options for part of speech and the second question had participants choose between three word meanings all within the chosen part of speech: 1 fully correct, 1 partially correct, and 1 incorrect (see Fig. 1).

Quantifying word salience We expected that people's ability to infer word meanings would be influenced by the fre-

[2]https://www.freechildrenstories.com/why-the-cricket-chirps

quency with which the target word occurred in the passage. But we also suspected that aside from frequency, performance would also be related to a word's *salience* (C. M. Brown, 1993). There is no single definition of salience, but intuitively, a word is salient to the extent that it communicates the central point of a story. For example, words naming the actions performed by a central character are more salient than words describing aspects of the environment that a non-central character inhabits. We quantified the salience of each target word as the likelihood that it would be recalled by people who read the original (unaltered) story. We recruited 46 participants from MTurk to read the original story and then had 1 minute to recall all the words they could remember occurring in the story. Salience for each word was defined as the sum of the weights that the word obtained each time it was listed: the weight was calculated as exponentially decreasing in accordance with the order in which words were listed by participants [(for each time the word was listed) weight = (0.75ˆ(word_n -1))] (see Table 1 for frequency and salience of each target word).

Table 1: Frequency, salience and morphological variation for target words

Target word	Frequency	Salience	Derivational Morphology	Inflectional Morphology
fly	18	6.59	1 derivational form (-er)	2 inflected forms (-ed; -ing)
cricket	16	31.69	none	1 inflected from (-s)
cold	7	5.69	none	none
wing	6	8.26	none	1 inflected from (-s)
fast	6	3.76	none	2 inflected forms (-er; -est)
chirp	5	8.49	none	3 inflected forms (-s; -ed; -ing)
ant	5	6.91	none	none
hop	5	1.06	none	none
snow	5	2.56	none	none
rub	4	1.90	none	2 inflected forms (-ed; -ing)
listen	4	1.47	none	none
warm	4	1.31	none	none
tree	3	0.44	none	1 inflected from (-s)
frozen	3	4.82	none	none
ground	2	0.00	none	none
owl	2	10.35	none	none

Procedure General procedure is shown in Figure 1. At the start of the task, participants completed a 10 item vocabulary test (Wordsum; Malhotra, Krosnick, & Haertel, 2007) and a pretest gauging familiarity with the target words. Participants were then randomly assigned to the real-passages or the nonsense-story condition. Those assigned to the real-passages condition saw each (meaningful) context and answered the two vocabulary questions for each of the 9 real-rare words (in random order). Each word was tested immediately after being presented in its context. The group was then tested on nonce words (skoast, etc.). In contrast, participants in the nonsense-story condition saw the entire 604-word nonsense-story and were then tested on the real-rare words, and then on 9 of the target nonce words (in random order). This design allowed each condition to serve as the control for the other condition. Subjects in the nonsense-story condition were asked to infer the meaning of the *real-rare* words without exposure to the passages and vice versa. As an attention check, scattered among nonce and *real-rare* words were questions about the meaning of familiar words (e.g. "little", "green").

Results and Discussion We analyzed the data using logistic mixed effects models. In the initial analyses, we treated partially and completely correct scores for specific meaning as the same (i.e., a binary contrast between an accuracy of 0 and 0.5/1). Figure 2 shows a clear interaction between condition (real-passages, nonsense story) and word-type (real-rare, nonce). This interaction was present both for part of speech [z = 7.3, SE = 0.28, $p < .001$] and word meanings [z = 9.1, SE = 0.29, $p < .001$] [accuracy ˜ type_of_word*condition+(1|subj_id)]. Participants in the real-passages condition were significantly more accurate at selecting meanings corresponding to the correct part-of-speech of real words compared to participants in the nonsense-story condition (i.e., those not exposed to the real passages) [z = -3.7, SE = 0.39, $p < .001$]. They were also better at choosing the more correct specific meanings [z = 4.6, SE = 0.4, $p < .001$] [accuracy (for specific meaning or for part of speech) ˜ condition+(1|subj_id)+(1|word)]. Note that above chance performance for the *real-rare* words is expected even without being exposed to real passages because some participants already know the meaning of these words. Not surprisingly, accuracy on *real-rare* words was positively associated with greater vocabulary as assessed by Wordsum. This was true both for part of speech [z = 2.8, SE = 0.17, p = .005] and specific meaning measures [z = 3.7, SE = 0.19, $p < .001$] [accuracy (for specific meaning or for part of speech) ˜ condition*wordsum_score+(1|subj_id)+(1|word)]. Greater familiarity with the target *real-rare* words (pretest) was positively related to selecting the correct part of speech [z = 2.1, SE = 0.16, p = .033] and specific meaning [z = 2.9, SE = 0.16, p = .004].

The results for the real-rare words tell us what we already knew – people can infer word meanings from seeing the words in context. We now turn to the nonsense-story condition. Recall that these nonce words were seen in the context of a 600+ word story in which *all* content words were replaced by nonce words. Participants exposed to this extremely sparse context had significantly higher performance in inferring the correct part of speech for the nonce words [z = 3.9, SE = 0.23, $p < .001$] and in choosing the more correct meanings [z = 4.4, SE = 0.25, $p < .001$]. The benefit from reading the nonsense story was not limited to just helping people choose the correct part of speech. Restricting the analysis to only the trials on which participants chose the correct part of speech, we find that exposure to the nonsense-story still led to higher accuracy [z = 2.9, SE = 0.35, p = .004] [$\text{accuracy}_{\text{part\_of\_speech}}$ ˜ condition+(1|subj_id)+(1|word)]. Neither Wordsum nor Pretest scores predicted performance for nonce words. Word frequency and word salience likewise did not predict performance ($z<1$, but see Exps. 2 and 3).

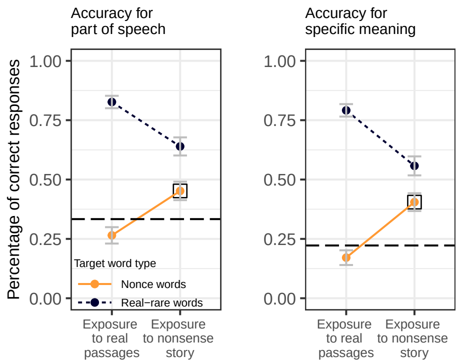

Figure 2: Accuracy for type of target word for Experiment 1. Horizontal dashed lines indicate chance-level. Error bars indicate +/-1SE of the mean. Significant effect of context exposure for nonce words in the nonsense-story is marked by a squared shape around the relevant data-point in the graph.

Experiment 2

Experiment 1 showed that people can benefit from very sparse contexts. One shortcoming of the study was that participants in the real-passages and nonsense-story conditions were exposed to contexts in a different way (short passages vs. one long story; tested after each word vs. tested on all words at once). In Experiment 2, we test participants in the same way on all the same words, varying just the informativeness of the context. We were also curious about whether it mattered how word knowledge was assessed. Instead of using the "drill-down" format that first asked about part of speech, we used a more standard multiple-choice text, presenting all the 9 options for each word at the same time. Lastly, we parametrically varied the informativeness of the context by replacing different proportions of content words with nonce words.

Methods Participants were randomly assigned to one of six conditions (see Table 2 for a summary). In the real-text condition we exposed participant to the original *Cricket* story, in which we replaced only the target words with *real-rare* English words. Thus, the linguistic context was still informative (i.e., the target words were surrounded by meaningful words) but it did not directly aim at communicating the meaning of the target words, e.g., *'You should find some shelter from the cold night,' said the smew. The mitius did not auscult, because mitiuses do whatever they want. He decided to rest on a pile of twigs.* We progressively decreased the information provided by the context by replacing various proportions of the remaining content words in the story with nonce words (40%, 60%, 90%, 100%). Lastly, we included a control group that was tested on their knowledge of the *real-rare* words without seeing any prior context.

Participants We recruited 246 participants from Amazon Mechanical Turk (112 Male of average age = 35; 132 Females of average age = 38). 38 participants were assigned to the real-text condition, 36 to the 40% condition, 39 to the 60% condition, 42 to the 90% condition, 52 to the 100% condition and 39 to the control condition.

Procedure Participants in each condition were initially tested on Wordsum and Pretest (following the procedure for Experiment 1) and then randomly assigned to one of the six conditions described above. All participants were then presented with the same 12 multiple choice questions (9 options per question) to assess their knowledge of the *real-rare* words.

Results and Discussion Results are shown in Fig. 3. Participants exposed to the full story clearly benefited in inferring both part of speech [z = 5.9, SE = 0.22, p < .001] and specific meaning [z = 5.6, SE = 0.25, p < .001] of *real-rare* words [accuracy˜overall ˜ control_vs_full_story +(1|subj_id)+(1|word)]. Similar results were found for the 40% and the 60% conditions, in which participants showed compelling effects for both part of speech [z = 4, SE = 0.22, p < .001] and specific meaning [z = 3.8, SE = 0.25, p < .001]. In contrast, when 90% or 100% of content words were replaced with nonce words, no significant benefit was observed for either part of speech [z = 1.2, SE = 0.2, p = .226] nor specific meaning [0.77, SE = 0.23, p = .444] results (Fig. 3).

Frequency of occurrence in the story was positively associated with accuracy for the real-text condition, [z = 2.7, SE = 0.16, p = .008] [accuracy ˜ condition*frequency+(1|subj_id)+(1|word)] and salience [z = 2.3, SE = 0.16, p = .023] [accuracy ˜ condition*salience+(1|subj_id)+(1|word)] in predicting accuracy for specific meaning. More frequent and more salient words benefited more from context. Similar effects were found in the 40% and 60% conditions for accuracy on specific meaning [frequency: z = 3.8, SE = 0.16, p < .001; salience: z = 2.3, SE = 0.16, p = .023] [accuracy ˜ condition*wordsum_scores+(1|subj_id)+(1|word)]. Controlling for pretest scores, greater vocabulary knowledge (Wordsum) was positively associated with accuracy for both part of speech [z = 2.7, SE = 0.12, p = .006] and specific meaning [z = 2.7, SE = 0.14, p = .007]. Similarly, previous word knowledge was positively associated with part of speech accuracy [z =2.4, SE = 0.084, p = .018] and specific meaning accuracy [z = 2.9, SE = 0.086, p = .003] [accuracy ˜ condition*pretest_scores+(1|subj_id)+(1|word)]. These associations parallel the findings of the real_rare words condition of Experiment 1.

Exposure to a story in which 40%-60% of content words were replaced with nonce words still allowed participants to learn something about meanings of words occurring in the story. There were two noteworthy differences between the results of Experiment 1 and 2. First, unlike Experiment 1, participants' ability to benefit from the story context was positively associated with the frequency with which the word occurred in the story and the word's salience. These relationships may stem from people's greater baseline knowledge of the (real-rare) target words. Second, exposure to a story in

Table 2: Summary of type of context, type of target word and methods is assessing word meaning in each experiment.

Experiment	Condition Name	Type of Text	Type of Target Word	Example Words	Question Test Type
Experiment 1	Real-passages	Real-rare target words; informative text	real-rare words	ratoon; pronk; rawky	drill-down
	Nonsense-story	Nonce-target words; nonsense story	nonce words	stronk; sprac; crex	drill-down
Experiment 2	Real-text	Real-rare target words: real story	real-rare words	auscult; lollop; smuir	multiple choice
	Real-words-nonce-context: 40%	Real-rare target words; 40% of context replaced with nonce words	real-rare words	auscult; lollop; smuir	multiple choice
	Real-words-nonce-context: 60%	Real-rare target words; 60% of context replaced with nonce words	real-rare words	auscult; lollop; smuir	multiple choice
	Real-words-nonce-context: 90%	Real-rare target words; 90% of context replaced with nonce words	real-rare words	auscult; lollop; smuir	multiple choice
	Real-words-nonce-context: 100%	Real-rare target words; 100% of context replaced with nonce words	real-rare words	auscult; lollop; smuir	multiple choice
	Control group	No exposure to story	real-rare words	auscult; lollop; smuir	multiple choice
Experiment 3	Real-words-nonce-context: 100%	Real-rare target words; 100% of context replaced with nonce words	real-rare words	auscult; lollop; smuir	drill-down
	Control group	Nonsense story with all words replaced with nonce words	real-rare words	auscult; lollop; smuir	drill-down

which all the content words with nonce words did not lead to greater-than-baseline performance on the word test. Experiment 3 was designed to better understand this difference.

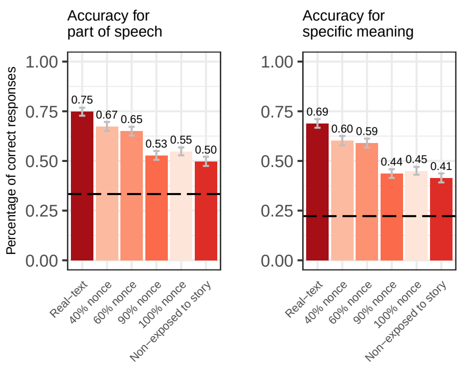

Figure 3: Group performance for Experiment 2. Horizontal dashed line indicated chance-level. Error bars indicate +/- 1SE of the mean.

Experiment 3

In Experiment 1, we found a significant effect of context exposure in inferring the meaning of nonce words from nonsense context. In Experiment 2 we assessed knowledge of *real-rare* words instead of nonce words, and embedded them in contexts of varying informativeness. Contexts in which 40-60% of content words replaced with nonce words were helpful, but those in which more (90%-100%) of content words were thus replaced, were not helpful. How do we reconcile this difference? Aside from testing nonce words vs. real-rare words, Experiments 1 and 2 differed in the way word knowledge was assessed. Experiment 1 first asked about part-of-speech. Experiment 2 presented all the meaning choices together, intermixing meanings from different parts of speech. We reasoned that explicitly asking people about parts of speech (which are more directly bootstrapped by morpho-syntactic cues) may make it easier for people to subsequently access more specific aspects of the word's meaning. In Experiment 3 we tested the effect of exposing people to a nonsense-story containing *real-rare* words as the *real-words-nonce-context (100%)* condition of Experiment 2, but using the drill-down question format of Experiment 1.

Methods Participants were randomly assigned to either the *real-words-nonce-context: 100%* or to a control condition. In the *real-words-nonce-context*, participants were exposed to a nonsense-story containing the *real-rare* target words and tested on those *real-rare* words (as in the 100% condition of Experiment 2). Participants in the control condition were shown a story with all nonce words (as in the nonsense-story condition of Experiment 1) but at test were asked about the meaning of the *real-rare* target words of the *real-words-nonce-context* condition (i.e., they were asked about words they did not see in the story).

Participants We recruited 81 participants from Amazon Mechanical Turk (39 Male of average age = 37; 41 Females of average age = 37). 41 to the *real-words-nonce-context: 100%* story condition and 40 to the control condition.

Procedure Participants in each condition were initially tested on Wordsum and Pretest (following the procedure for Experiment 1) and then randomly assigned to either the *nonsense-story* condition or to a control group that was not exposed to a story. All participants were tested on the same set of drill-down questions.

Results and Discussion Results are shown in Fig. 4. We found a significant effect of exposure to the linguistic context in inferring the part of speech [z = 3, SE = 0.18, p = .002] and the specific meaning [z = 2.6, SE = 0.2, p = .008] of *real-rare* words when compared with the control condition [$accuracy_{overall}$ ~ condition+(1|subj_id)] (Fig. 4). When examining only trials on which participants inferred the correct part of speech, the benefit of exposure to a nonce story on inferring the correct specific meaning was no longer significant [z = 0.5, SE = 0.31, p = .614] [$accuracy_{part\_of\_speech}$ ~ condition+(1|subj_id)+(1|word)]. To determine if the nonce-word context in the present study was more effective than in the equivalent condition of Experiment 2, we examined the interaction between experiment (*Exp. 2* vs. *Exp. 3*) and condition (*control group* vs. *real-words-nonce-context: 100%*) [accuracy ~ condition*experiment+(1|subj_id)+(1|word)]. This interaction was significant for both part-of-speech [z = 2.9, SE = 0.18, p = .003] and specific meaning [z = 2.5, SE = 0.21, p = .003]. As in Experiment 2, we found a significant effect of both frequency and salience of the target words. The benefit of being exposed to the nonce story was greater for more frequent words [for part of speech: z = 2.1, SE = 0.14, p = .035; for specific meaning: z = 2.5, SE = 0.15, p = .012] and more salient words [for part of speech: z = 2.2, SE = 0.14, p = .031; for specific meaning: z = 2.6, SE = 0.16, p = .009].

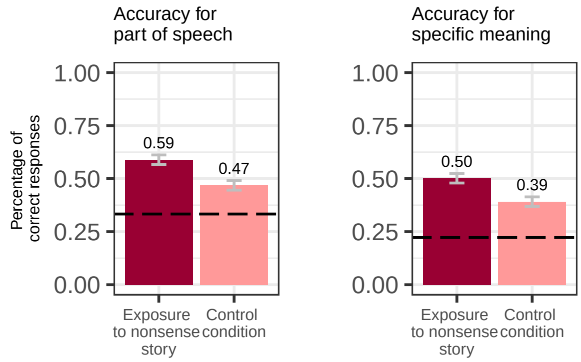

Figure 4: Group performance for Experiment 3. Horizontal dashed lines indicate chance-level. Error bars indicate +/-1SE of the mean.

Controlling for pretest scores, greater vocabulary knowledge (Wordsum) predicted higher accuracy for both part of speech [z = 3.4, SE = 0.091, p = .001] and specific meaning [z = 3.1, SE = 0.1, p = .002].

In Experiment 3, we observed an effect of context similar to that observed in Experiment 1. People were able to learn something about unknown words (here, *real-rare* words) from contexts in which all content words were replaced by nonce words. The only difference between Experiment 2 and the present study was in how word knowledge was assessed. A plausible conclusion – though in need of further testing – is that explicitly asking participants about a word's part of speech helps them deploy a more informative prior within which to consider the specific meanings.

General Discussion

What are the limits of learning word meanings from language? Our results show that participants were able to learn something about what a word means from brief exposures to such seemingly meaningless contexts as "The stronk rourthed daft to a dweave luk as the slom zeuded rhiecing." For example, after reading a 600+ word nonsense-word story containing sentences like the one just used, 38% of participants correctly inferred that "stronk" means "an insect" compared to 0% of participants who were not exposed to the story.

Our attempt to replicate and extend the results of Experiment 1 to using sparse contexts to inform the meanings of real, but rare and generally not known words (e.g., auscult, mitius), revealed that while partially informative contexts (40-60% of content words replaced with nonce words) were helpful, more sparse contexts (90%-100% of content words replaced) were not. We hypothesized that a key difference was the way that word knowledge was tested. In Experiment 1, participants' word knowledge was tested using a drill-down format that asked participants to first consider the word's part of speech. In Experiment 2, participants were asked to choose from among all the nine options visible at the same time making part of speech a less salient dimension of the word's meaning in the testing phase. In Experiment 3 we used the methods and materials of Experiment 2 with the test format of Experiment 1. Highlighting part-of-speech information using "drill-down" questions once again revealed that participants were able to use the nonce-story context to infer meanings of novel words.

Experiments 2-3 further showed that the effects of context were positively associated with frequency and salience. Words that were more frequent and more salient benefited more from context. While frequency is perhaps the most often used predictor in studies of word learning, to our knowledge, we are the first to examine the role of *salience*, defined here a the likelihood that people recall reading the word (see Table 1). What precisely makes a word salient requires future research.

What information did people use to infer word meanings? In the all-nonce-word conditions of Experiments 1 and 3, greater than baseline performance cannot be explained by reliance on the meaning of English content words because no recognizable content words were present. There were three remaining sources of information: closed-class words, inflectional cues, and syntactic cues. Consider one sentence from the story: "He thecked up into a dweave luk to fruth in for a sparf snurv." The remaining pronouns and prepositions combined with inflectional cues can clearly be used to infer that, e.g., "thecked" is an action being performed by an animate agent and that a "dweave" is likely to be some kind of place. Implicit knowledge of English syntax such that verbs follow "to" and objects tend to come after verbs offers further guidance. What is remarkable is that participants are making these inferences in parallel across dozens or even hundreds of words and that a single exposure to the story is sufficient to achieve above baseline accuracies.

Our work has two main limitations. First, successful use of sparse contexts involving mostly or exclusively nonce words clearly requires participants to already have sophisticated knowledge of English and so while it can help us understand how adults learn new words from context, it does not tell us how people learn enough English to make use of such sparse contexts. Second, present experiments do not tell us the relative importance of closed-class words, syntax, and morphology. Answering this question would require manipulating these sources of information independently. We can also gain additional insights by conducting studies such as this in more morphologically rich languages.

Conclusion

As has been long known, people are able to learn something about a word's meaning from encountering it in context. What is surprising is just how sparse and seemingly uninformative that context can be. The facility that people show in inferring word meanings from such sparse and seemingly meaningless contexts suggests that we may be underestimating the role that morpho-syntactic and distributional cues have on both learning word meaning and on acquiring semantic knowledge that is embedded in language.

Acknowledgements

This work was partially supported by NSF-PAC 1734260 to GL.

References

Brown, C. M. (1993). Factors affecting the acquisition of vocabulary: Frequency and saliency of words. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 263–286). Norwood, NJ: Ablex.

Brown, R. W. (1957). Linguistic determinism and the part of speech. *The Journal of Abnormal and Social Psychology*, *55*(1), 1–5.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911.

Firth, J. R. (1957). A synopsis in linguistic theory, 1930-1955. In.

Gleitman, L. (1990). The Structural Sources of Verb Meanings. *Language Acquisition*, *1*(1), 3–55.

Goldberg, A. E. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences*, *7*(5), 219–224.

Ingraham, A. (1903). *Swain school lectures*. Open Court Publishing Company.

Landauer, T. K., & Dumais, S. T. (1997). A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge. *Psychological Review*, *104*(2), 30.

Lany, J., & Saffran, J. R. (2010). From Statistics to Meaning: Infants' Acquisition of Lexical Categories. *Psychological Science*, *21*(2), 284–291.

Levy, O., & Goldberg, Y. (2014). Neural word embedding as implicit matrix factorization. In *Advances in neural information processing systems* (pp. 2177–2185).

Malhotra, N., Krosnick, J. A., & Haertel, E. (2007). The psychometric properties of the gss wordsum vocabulary test. *GSS Methodological Report*, *11*.

McDonald, S. A., & Ramscar, M. (2001). Testing the Distributional Hypothesis: The Influence of Context on Judgements of Semantic Similarity. *Proceedings of the Annual Meeting of the Cognitive Science Society, 23(23)*, 7.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space.

Muckenhoupt, C. (2001). The Gostak.

Nagy, W. E., & Anderson, R. C. (1984). How Many Words Are There in Printed School English? *Reading Research Quarterly*, *19*(3), 304.

Nagy, W. E., & Herman, P. A. (1984). Limitations of vocabulary instruction (Tech. Rep. No. 326). *Urbana: University of Illinois, Center for the Study of Reading*, 30.

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning Words from Context. *Reading Research Quarterly*, *20*(2), 233.

Naigles, L. R. (1990). Children use syntax to learn verb meanings. *Journal of Child Language*, *17*(02), 357.

Naigles, L. R., & Swensen, L. D. (2007). Syntactic Supports for Word Learning. In E. Hoff & M. Shatz (Eds.), *Blackwell Handbook of Language Development* (pp. 212–231). Oxford, UK: Blackwell Publishing Ltd.

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC Nonword Database. *The Quarterly Journal of Experimental Psychology Section A*, *55*(4), 1339–1362.

Redington, M., Chater, N., & Finch, S. (1998). Distributional Information: A Powerful Cue for Acquiring Syntactic Categories. *Cognitive Science*, *22*(4), 425–469.