Noun Bias in Adulthood Found to Depend on Test Delay and Learning Method

Jason Darryl Ludington

Abstract

Among adults learning a foreign language, might nouns be easier to learn than verbs? A noun learning bias is characteristic of child learners across most languages (e.g., Gentner, 1982), but few studies have tested for such a word class learning bias among adults learning a foreign language. To test adult susceptible to a noun bias, I trained 84 participants with novel names of 96 familiar, concrete concepts (half nouns, half verbs) by ostensive and inferential training methods, and tested vocabulary word recognition at five minutes and at one week. I tested a number of stimulus features which literature suggested could be of importance to word learnability (23 factors in total). This led to the development of a controlled test of the influence of word class apart from other contributing factors on learnability. There appears little evidence of a generalized noun bias in adulthood—a noun bias was found but was specific to ostensive training, tested five minutes but not one week later. Word imageability and image quality also contributed to vocabulary meaning recognition. Results support a theory of word learning that emphasizes referential clarity. Also discussed is the possibility of recognition likelihood as a dynamic construction.

Key words: noun learning bias, adult L2 learning, vocabulary meaning recognition

บทคัดย่อ

เป็นไปได้หรือไม่ว่าสำหรับผู้ใหญ่ที่เรียนภาษาต่างประเทศนั้นการเรียนรู้คำนามมักจะง่ายกว่าการเรียนรู้คำกริยา อันที่จริงแล้ว แนวโน้มที่จะเรียนรู้คำนามก่อนหน่วยไวยากรณ์อื่น ถือเป็นลักษณะเด่นของผู้เรียนเยาว์วัย ที่เรียนภาษาเกือบทุกภาษา (เก็นเนอร์ 1982) แต่มีงานวิจัยน้อยชิ้นมากที่ทดสอบแนวโน้มของผู้เรียนภาษาต่างประเทศที่เป็นผู้ใหญ่ในการเรียนหน่วยไวยากรณ์ เช่น คำนามได้เร็วกว่าหน่วยไวยากรณ์อื่น ๆ และ

เพื่อเป็นการทดสอบแนวโน้มในการเรียนรู้ คำนามในผู้ใหญ่นี้ ผู้วิจัยได้สอนคำต่าง ๆ จำนวน 96 คำ ที่ล้วนมีความหมายเป็นรูปธรรม (แบ่งเป็นคำนามครึ่งหนึ่งและคำกริยาครึ่งหนึ่ง) ให้แก่กลุ่มตัวอย่างจำนวน 84 คน ซึ่งไม่เคยรู้จักคำเหล่านี้มาก่อน โดยผู้วิจัยใช้วิธีการสอนแบบชี้ชัดและแบบอ้างอิง ตามด้วยการทดสอบการรู้คำศัพท์ในช่วงระยะเวลา 5 นาที และหนึ่งสัปดาห์หลังการสอน ผู้วิจัยยังได้ทดสอบลักษณะของตัวกระตุ้น ซึ่งงานวิจัยชิ้นก่อน ๆ เสนอแนะว่ามีความสำคัญยิ่งต่อความสามารถในการเรียนรู้คำศัพท์ (รวมทั้งสิ้นเป็นจำนวน 23 ปัจจัย) การทดสอบดังกล่าว นำไปสู่การสร้างแบบทดสอบเชิงควบคุม เกี่ยวกับอิทธิพลของประเภทคำนอกเหนือไปจากปัจจัย สาเหตุอื่น ๆ ที่มีผลต่อความสามารถในการเรียนรู้ ผลวิจัยพบว่ามีหลักฐานสนับสนุนน้อยมาก เกี่ยวกับแนวโน้มในการเรียนรู้คำนามของผู้เรียนวัยผู้ใหญ่ กล่าวคือ พบว่ามีแนวโน้มในการเรียนรู้คำนามเฉพาะในกรณีของการสอนแบบชี้ชัด และเมื่อมีการทดสอบ 5 นาทีหลังจากนั้น แต่ไม่ปรากฏผลดังกล่าวเมื่อทดสอบหนึ่งสัปดาห์ต่อมา นอกจากนี้ ยังพบว่า ความสามารถในการจินตนาการคำศัพท์และคุณสมบัติของภาพที่จินตนาการมีผลต่อความสามารถในการรับรู้ความหมายของคำศัพท์ ผลวิจัยนี้สนับสนุนทฤษฎีการเรียนรู้คำศัพท์ที่เน้นความชัดเจนของความหมายที่อ้างถึง และยังได้อภิปรายถึงความเป็นไปได้ของแนวโน้มในการรับรู้ในฐานะที่เป็นโครงสร้างเชิงพลวัตร

คำสำคัญ : แนวโน้มในการเรียนรู้คำนาม การเรียนรู้ภาษาที่สองในผู้ใหญ่ การรับรู้ความหมายคำศัพท์

Introduction and Overview

My experiences teaching English as a second language in universities in Thailand impressed upon me how valued and vital second language teaching and learning is across the globe. This interest brought me to the literature on vocabulary acquisition, the noun bias debate, and methods of vocabulary learning, topics addressed in this paper.

The importance content words when learning a new language

When infants initially speak their mother tongue, they enter a one-word stage followed by a two-word stage of language development (Greenfield & Smith, 1976). The kinds of words they first use are not random—they are virtually always frequently used, concrete, content words, in particular mostly nouns and verbs (e.g., “mommy,” and “up,” as in “pick me up”). It is not by mistake that infants speak these kinds of content words first: nouns and verbs convey meaning in and of themselves, unlike function words (e.g., “of” and “a,”), and are therefore among infants’ first communicative acts. More aged learners may not be all that different. Krashen and Scarcella theorized adults and children are probably similar with regard to the way that language acquisition proceeds from one-word-at-a-time to more complex usage, if not for different reasons (1981, p.296). It would be impractical and improbable for any learners, regardless of age, to begin speaking in full sentences or to begin using infrequent, abstract “higher level” vocabulary words. Where the primary purpose of speaking is to communicate meaning, word production is likely to begin with short utterances of concrete content words, as they communicate meaning even as incomplete phrases.

Most content words fit neatly into the categories noun, verb, adjective and adverb, with the latter two as descriptors of the former two. This study focuses only on former two, and focuses exclusively on mature learners where there is a much greater poverty of research on learning differences between word classes.

Two methods of acquiring vocabulary

Building one’s vocabulary is a good way to begin learning a new language, but how should this be done? Implied in most language textbooks and language teaching methods is the notion that vocabulary is best learned by ostensive labeling of a concept (translation would fit into this learning category). This seems to be the most obvious way to bring awareness of word meaning to the surface. However another, less direct method is to let learners discover meaning by themselves from a meaning-rich context. The implication with this choice of teaching method is that the process of discovering or inferring is, itself, acting to deepen the encoding of target material.

I manipulated method of learning at two levels. The direct method is called “ostensive” learning and the indirect method “inferential” learning. Ostensive and inferential learning conditions are illustrated in Figure 1. In ostensive learning trials, a pair of labels (a noun and a verb) was learned from individual presentations of isolated examples, called “isolate images.” Inferential conditions were created so that participants would have to infer the meaning of one of two elements presented in a context image. In inferential conditions, a single isolate image and a “context image” were shown, in that order, which allowed learners to recognized the redundant element and thus infer the meaning of one of the two target elements from their context.

Ostensive, one occurrence per trial

“yainoop” “jev” “yainoop jev”

Ostensive, two occurrences per trial
“yainoop” “jev”

Inferential, one occurrence per trial
“yainoop”
“yainoop” “jev”

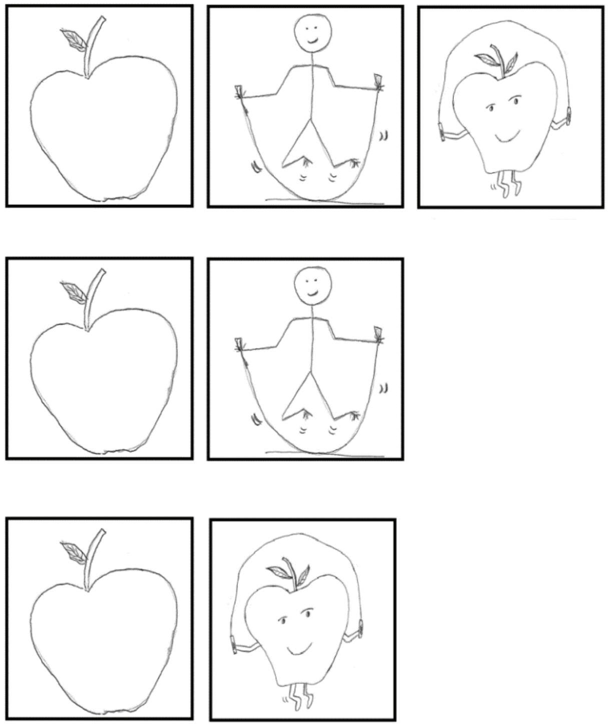

Figure 1. Flow diagrams illustrating ostensive (top and middle panels) and inferential learning (bottom panels) of the verb “jev,” meaning “to jump rope”. To avoid copyright violations these images were created by me and an anonymous colleague (actual experiment images are not shown).

Notice in Figure 1 that in the inferential condition (shown in the lower panel in the figure), a learner should gather that “jev” means “to jump rope.” The learner can infer this based on the principle of mutual exclusivity (Markman & Wachtel, 1988), an application of the following four-step logic: 1) “yainoop” is known, based on the first image, to refer to the apple; 2) “yainoop” cannot refer to jumping rope because it refers to the apple; 3) “jev” should not refer to apple because “yainoop” refers to apple; therefore 4) “jev” must refer to jumping rope because that is an otherwise unnamed referent in the last image.

The noun bias debate

Gentner’s (1982) seminal research finding was of a noun bias pervading the six languages she studied. This led to further research that has found children across most languages and cultures acquire nouns faster than verbs. Verbs, it seems, are the more difficult of the two word classes for young learners to learn. While a great deal of research has focused on why some words are harder than others for children to learn, much less research has addressed what makes words easier or harder for adults to learn in a second or subsequent language. Does the noun bias reign on into adulthood, and if so, why?

Many theories have been proposed to explain the noun bias in early word learning: natural partitions and relational relativity (Gentner, 1982), location of nouns within utterances (Shady & Gerken, 1999; Au, Dapretto, & Song, 1994; Tardif, 1996; Tardif, Shatz, & Naigles, 1997), verb argument requirements (Greenfield & Alvarez, 1980; Gleitman, Cassidy, Nappa, Papafragou, & Trueswell, 2006; Gentner, 2006; Waxman & Lidz, 2006; Sandhofer & Smith, 2007), inherent verb complexity (Akhtar, Jipson, and Callanan, 2001; Tomasello, 1992), syntactic complexity (Pinker, 1994; Naigles, 1990), cultural emphasis (Gopnik & Choi, 1990, 1995), attention allocation and category membership variability (Kersten, Smith, & Yoshida, 2006), to name a few. In this paper I propose that it is the referential clarity with which meanings are presented in learning contexts that can explain why some words are harder to learn than others, regardless of the learner's age.

Evidence suggesting a noun bias in adulthood

Gillette, Gleitman, Gleitman, & Lederer (1999) reported on the results of their "Human Simulation Paradigm" in which adults were shown video interactions between mothers and their infants. The videos were silenced, and beeps were inserted in places where the mothers uttered mystery words. The test was to see whether adults might guess the mystery words based on what they could see from the video. Only the earliest acquired nouns and verbs by children were sampled for experimentation. Gillete et al. found that adults were better at guessing noun meanings than verb meanings. They also showed the benefits of noun knowledge for guessing verbs, of syntactic frame knowledge, and of combined information knowledge for guessing word meanings. Besides making a powerful statement for the roles of context knowledge and grammar in learning new vocabulary, their study showed evidence of a noun bias in adulthood. They found that concreteness, which was correlated with lexical class, was the underlying factor that allowed nouns to be learned quicker than verbs. This noun advantage was replicated in Gleitman and colleagues (2006) word learning experiment findings, as well as by Snedeker & Gleitman (2004).

Referential Clarity Hypothesis

Verbs may be harder to acquire because they are harder to point at (Greenfield & Alvarez, 1980), or because there are fewer constraints on their referential intention (Gentner, 1982). These two views seem in conflict but they were developed under different learner conditions. While the latter theory (Gentner) does very well to explain the noun bias among those learning their first language, the former (Greenfield & Alvarez) may apply equally well to more aged learners of a second or subsequent language.

Greenfield and Alvarez (1980) taught words by uttering sentences and presenting images to illustrate sentence meanings. They found that as the number of unknown words and meanings in a context decreased, vocabulary learning increased. Having fewer unknown words in the learning context helped to clarify the word mapping problem and thus enhanced learning. The referential clarity (RC) hypothesis, based on that study's findings, is that word learning difficulty is a matter of situational or contextual ambiguity contained in word learning situations; thus the greater the referential clarity of a given word learning situation, the more learnable is that word in that situation. Verb learning situations always arise with actors, and sometimes with patients and tools of enactment; these potential learning situations tend to arise with too much referential ambiguity to allow learners to effectively map verbs to their meanings. Prior exposure to these contextual arguments can aid verb learning by disambiguating which of multiple possible referents correctly maps to the verb. Noun knowledge can function to reduce ambiguity of verbal referents. Verb knowledge, however, may not be as helpful to noun learning because nouns are typically less ambiguous to begin with.

What I did

I presented college students with 96 foreign words, half of which stood for concrete nouns and half concrete verbs. Given the past research suggesting a noun bias in adulthood I hypothesized I would also find a noun advantage. Besides testing the effect of word class on "word learnability" or likelihood of a word being learned from exposure in a controlled setting, I also tested several other potential indicators of word learnability, and conditions that might lead to differential learning success. High on this priority list was measuring a hypothesized advantage of inferring word meanings rather than having them simply pointed out. The characteristics of my stimuli that I tested were measured in the unpublished work of a prior study; English words, their images, and nonsense words were measured to establish their role as predictors, and to control them if necessary, in assessing the evidence for a noun bias.

I tested word knowledge by testing word recognition performance rather than a recall test because cued recall tests may be too insensitive to the kind of under-developed knowledge forms that would come from learning novel words from limited examples within limited time periods in the present paradigm (i.e., I wanted to avoid floor effects). In this paper I do not make a distinction between word recognition and word learning. This view is defensible: in cognitive psychology recognition and recall are frequently used as analogs of one another to assess subjects' states of knowledge. Recognition and recall have long been conceptualized as two features of the same construct—memory. Testing recognition rather than free recall is much like using a stethoscope rather than an EKG monitor to measure someone's pulse—though both measure the same construct, they do so by measuring different biophysical phenomena.

Predicted main effects

I predicted a noun bias, as found in Gillette et al. (1999) and other studies. However if my participants learned more of the noun than verb targets, one could argue the verbs were harder to learn than the nouns for other reasons besides word class. In anticipation of this I also documented many other features of the target stimuli to test them as explanations of a noun bias if found, and of word learnability in general.

If the difficulty of meaning identification in the environment is to blame for poor vocabulary acquisition (i.e., the RC hypothesis), then name agreement measurement of visual stimuli (the proportion of naming responses fitting pre-established target responses) could be a powerful predictor of word learnability with those images. In the present study I used images of line drawings to model real world referents; name agreement measures were known for these line drawings as they were measured as part of a prior study. I predicted that name agreement measurements, as measurements of image clarity, would account for any word class bias I might find; this prediction was based on the RC hypothesis, the idea that vocabulary learnability is as a linear function of referential clarity. By proposing this hypothesis as an account of word learning, I wish not to insinuate this is the only method by which learners acquire words; certainly grammatical and statistical learning cues and other documented methods of learning greatly contribute to vocabulary acquisition. I only wish to propose that when a target label's meaning is present in the learning environment at the time it is uttered, the RC hypothesis may account for a major portion of the learning that follows.

Past research suggests people typically rate noun images as better depictions of their intended referents than verb images (Kauschke & Frankenberg, 2008; Masterson & Druks, 1998). This measure, in addition to name agreement measures of images, should quantify how clear or ambiguous meanings are “given” by their images. Alternative interpretations of images similarly could be a good indicator of meaning clarity in images. I predicted that all three measures of image quality could attenuate or account for any learning disparity I might find between nouns and verbs. I predicted that word imageability (Gillette et al., 1999) also could account for any word class effects found. These predictions are congruent with the RC hypothesis because measures of image clarity are measures of referential clarity, and word imageability could affect the “clarity” or likelihood of identifying meanings presented in images.

Subjective ratings of concept frequency (the frequency with people encounter the target concepts in their lives) and word familiarity were also considered as possible predictors of word learning performance. Rated word imageability is usually correlated with familiarity and frequency (e.g., Stadthagen-Gonzolez & Davis, 2006), however, which might present a challenge for determination of which of these correlates better accounts for outcomes.

Predicted interactions

An important purpose of this study was to investigate whether inferential learning was better suited for learning verbs than nouns. Concrete noun vocabulary can be learned by labeling of that which is referred to, attended to, and pointed at. Because verbs arguably cannot be pointed at (only their arguments can be pointed at—actors, patients, tools, etc.) ostensive training (labeling; e.g., “this is ____”) may not be available, and thus verb acquisition may rely on inferential processes instead. As adults learning a second language approach their goal with a great deal of prior verb learning experience (from learning their native language), they may be more skilled at inferring verbs than nouns. It was thus my prediction that ostensive learning would lend better to learning nouns, but inferential learning to verbs.

Another important purpose of this study was to address a question regarding the effect of delay on recognition. I measured learning at two learning delays—five minutes and one week. By measuring twice, I could test for a forgetting differential between ostensive and inferential learning conditions. Most fast-mapping studies test participants within a matter of seconds or minutes, but at such short delays it is questionable whether fast-mapped words are truly learned or were just temporarily held in working memory. A few other studies have demonstrated that fast-mapped learning does remain measureable after longer delays, anywhere from five minutes (Horst & Samuelson, 2008), one or two days (Jaswal & Markman, 2003, experiment 1; Woodward, Markman, & Fitzsimmons, 1994), to one week or longer (Carey & Bartlett, 1978; Vespoor & Lowie, 2003), but very few studies have manipulated learning methods at multiple delays.

Knight (1994) is one notable exception: Knight found that ostensive learning (by glossing with a dictionary) aided learning more than inferential learning (guessing meanings from a reading, followed by confirming or disconfirming by glossing) on an immediate test and on a one-week delayed test. But Knight's inferential condition was not purely inferential; it was really inferential and ostensive learning together, and it took longer than either method by itself.

I wanted to test the effect of delay to see if inferential learning reduced forgetting, so I manipulated test delay at two levels, five minutes and one week. After five minutes I predicted participants would recognize more words learned ostensively than inferentially, as Knight found, because inferential learning necessarily involves some uncertainty of mapping words to their referents, while ostensive learning involves greater certainty of associations. But after one week I expected better recognition of words learned inferentially than ostensively due to slower forgetting of inferentially learned words. Inferential learning seems to require more processing than ostensive recognition because inferential learning requires hypothesizing followed by confirmatory testing. More involved or effortful processing is typically associated with longer term retrieve-ability (Bjork, 1994). As a deeper level of processing, I predicted inferentially learned words would be better recognized than ostensively learned words (a hypothetically shallow form of learning) at one week.

Other variables tested

Sources linking variables measured in the present study to learnability were not always available; in some cases a speculative leap was made in predicting their effects on learnability from their known effects on other performance measures (e.g., reading speed). I measured word familiarity which has been linked to faster reading speed (Brown & Watson, 1987), and word imageability which has been linked to lexical decision speed (Balota, Yap, & Cortese, 2006), and concept frequency. Category representation is a variable I measured to tap the von Restorff effect (1933), the well-established finding that unique items in a list stand out in memory. Category representation was measured as the percentage of targets presented, within a given half of the experiment (or “list”) that were members of each envisaged category.

I measured name agreement for images of targets in isolation and images of targets in context**,** as name agreement has been shown to affect naming speed (Ellis & Morrison, 1998). I measured ratings of how well the images conveyed their intended meanings, and number of alternative interpretations (raw total number of alternative responses offered by participants who provided this data), for which non-primary responses have been associated with longer response latencies (Székely et al., 2003). I measured ratings of how strange each given noun-verb pair was because I thought strangeness could positively affect memory performance. Finally, I measured auditory stimulus lengths in terms of utterance time, number of phonemes and number of syllables to see if word length might affect word learning. All of the measures described in this paragraph were obtained from a sample of 29 participants (who did not participate in the present study’s experiment) from my own unpublished work.

Partway through data collection I recognized the need for an additional manipulation of ostensive learning conditions. Ostensive conditions were initially created with a redundant label occurrence, but inferential conditions were created without redundant labeling. Therefore to separate the effect of number of occurrences from the effect of the method of learning, I ran an additional experimental condition with additional participants in which words were presented ostensively without any redundancy. Procedural differences between the initially-begun experiment and the added condition were minor enough to warrant inclusion of both conditions under the same experimental name. Thus “number of occurrences” was also measured in this study.

The above-described research questions are summarized below.

1. Is there a noun bias among adults learning foreign vocabulary?
2. Does the method of learning matter? Does inferential learning slow forgetting?
3. Is one method of learning better suited to learning verbs, and the other to nouns?
4. What are some other predictors of word learning?

Method

Participants

Ninety participants from the University of California, Los Angeles were recruited with an online recruitment site as used in the previous study. Some participants were dropped due to participants' failure to return for the second part of the experiment (5) or for experimenter failure to present all materials (1). The mean age of the remaining 84 participants was 20.8 years, SD=4.19 years. More females (62; 73.8% of sample) than males (22) participated. Most participants were at least partially able to use a second language (only 2 did not report any second language ability). The average number of languages (including English) reported at any proficiency on a 1–10 scale was 3.5. Participants reported their proficiencies in all languages including English on a 1-10 scale. The average language proficiency sum across participants and across all languages besides English, based on the aforementioned scale, was 7.8 (SD=4.5).

Design

Twenty-three predictors were tested in total, about half of them continuous, and half categorical. Some were experimentally manipulated, and some were not. Most of the categorical factors were manipulated within participants. Word class (nouns versus verbs), method of learning (ostensive versus inferential), and number of word lists learned prior (none versus one) were all manipulated within subjects. Experiment languages (Language A vs. Language B) and learning schedules (schedule 1 versus 2; these will be defined in the next section) were manipulated between subjects. Number of occurrences per trial was manipulated with subjects for 56 participants, but only at one occurrence among 28 participants. Delay was manipulated at five minutes and one week, within-subjects, except among the aforementioned 28 participants for whom delay was fixed at five minutes. Word order (order with which isolate images were presented: noun-verb versus verb-noun) was counterbalanced and manipulated within participants. Word order and number of occurrences were analyzed separately because both were variables nested within the ostensive level of the method of learning. Participants' sex and English-as-first-language status (English as most proficient language versus not) measured among the present sample of participants as potential predictors of learning.

There were a number of continuous variables tested as well: participants' age, self-rated English proficiency (scale of 1–10), and other language proficiency sum (the sum of all self-reported, self-rated proficiencies in languages other than English) were measured and tested in the present study. Additionally the following stimulus measurements, garnered from a prior (unpublished) study I ran with 29 participants, were tested: category representation, word imageability, word familiarity, concept frequency, goodness of depiction, number of alternative interpretations, word-pair strangeness, name agreement in isolate images as well as context images, and utterance lengths measured as number of phonemes, syllables, and time of utterance (in seconds and hundredths of seconds).

The dependent variable was the correctness of each target word selection made (i.e., recognition), measured on a binary scale as correct versus incorrect. A 25% likelihood of target selection marked chance performance as the recognition task was to select the correct target from among four targets. This outcome variable, measured for each word learned, was a measure nested within each participant (i.e., each participant was measured multiple times). Thus it was the word unit, nested within the participant unit, which was analyzed.

Materials

A consent form and biographic data form were used to collect hand-written data. The biographic data sheet collected participants' age, sex, self-reported proficiency in English, and self-reported proficiencies in other languages reported as known. SuperLab (stimulus presentation software) and a Toshiba laptop computer (16:9 LCD display) were used to present words and images for learning as well as testing. Following are descriptions of each measurement type.

Biographical data form.

This was used primarily to collect language background information. One question addressed what the participants' first language was. If not English, another question asked participants to rate their language ability in English on a fluency scale from 1-10, where 1 = unable to use any of the language, and 10 = fluent. A third question asked for other languages the participant knew, and how fluent he or she was in each (using this same fluency scale). Age and sex data were also collected.

Variables.

From a prior-run, unpublished study, measures of experimental targets' English word meanings (familiarity, frequency, imageability) and measures of image stimuli (isolate image name agreement, context image name agreement, goodness of representation, number of alternative interpretations offered) were measured using a separate sample of 29 participants (undergraduates from the same university, recruited in the same way). Table 1, below, gives this data.

Table 1 *Sample Ns, Means, and Standard Deviations*

			Nouns		Verbs	
Factor	Scale	N	Mean	SD	Mean	SD
Familiarity	1 to 7	26	6.45	0.84	6.49	0.62
Frequency	1 to 5	20	3.87	0.43	3.43	0.55
Imageability	1 to 7	26	6.67	0.48	6.68	0.35
Goodness	1 to 5	20	4.91	0.09	4.75	0.22
Alternatives	raw #	20	3.77	1.98	5.6	3.2
Isolate Naming	0 to 1	19	0.90	0.13	0.86	0.11
Context Naming	0 to 1	10	0.90	0.07	0.78	0.15

The variables, explored in the present study as potential predictors of word recognition, are listed in Table 2. Given the number of factors considered, I relate details of coding, measurement, analytical procedures and results together by variable for a simpler organization of information.

Table 2 *All 23 Factors Explored in Study 2*

	Item Factors	Participant Factors
concepts	Familiarity	Age
concepts	Frequency	Sex
concepts	Imageability	English-as-1st-lang status
concepts	Category representation	English proficiency
Images	Name agree, isolate	Language proficiency sum
Images	Name agree, context	
Images	Goodness of image	<u>Condition Factors</u>
Images	Alternative interpretations	Method of learning
Images	Strangeness of context	Word class
	Utterance-length	Word order
		Delay until test
		Experiment language
		Lists learned prior
		Occurrences/trial
		Learning schedule

Nonsense words.

Auditory stimuli were created, 96 nonsense words in total, 48 randomly assigned as noun labels, and 48 as verb labels. I decided not to use an existing language because real languages contain a mix of familiar and unfamiliar phonemes and phonemic structures, variables that I wanted controlled. Instead I created a mix of one- and two-syllable words to simulate words of a real language.

Nonsense words were created using a pool of 17 consonant phonemes and 7 vowel phonemes following a CVC (one syllable) or CVCVC (two syllables) structure to simulate words of a real language. Table 3 lists the consonants and their position rules. The set of consonants for word-initial and word-final positions was made partially overlapping, as is often the case in real languages. In many languages, some phonemes never or rarely occur in initial or final locations (e.g., "ng__ in English, and "__r" in Thai). Nonsense words were largely adopted from Vitevitch and Luce (1999) or adapted from the same phonemes they used to make their nonsense words, which were constructed from highly common in English phonemes. Highly familiar phonemes are more perceivable than unfamiliar phonemes (Appleman & Mayzner, 1981), though they may not be any easier to remember. Balanced numbers of one- and two-syllable words were created and assigned as nouns and verbs, 36 one-syllable and 12 two-syllable words for each word class.

Table 3
Phonemes Used to Construct Nonsense Word Stimuli

Location in Syllable	Phonemes
Beginning	D, F, G, H, J, K, L, N, P, R, S, Sh, T, Th, W, Y, Z
Middle	Ai, Ee, Eh, Ir, O, Oo, Uh
Ending	B, Ch, D, F, G, H, Jsh, K, L, M, N, P, S, T, Th, V, Z

Nonsense words were spoken by a native English-speaking Caucasian adult male (me) and recorded using Audacity 1.3.12 (Beta) (a free sound recording software) which was also used to edit and measure utterance lengths of all auditory stimuli. Sounds were edited to include a 100 milliseconds onset delay so that they would not be sounded simultaneously with image onset (to avoid distraction or reduced attention to either sense modality, either sight or hearing, when presented together). Individual words were recorded for presentation with isolate images, and two-word phrases were recorded for presentation with context images. I spoke and recorded phrases with normal sentential intonations (as a continuous utterance, not staccato words) to maintain the ecology of stimuli as complete phrases. Word utterance lengths, measured to the nearest hundredth of a second, were submitted to an independent samples t-test to check whether nonsense words assigned to one word class or the other systematically differed in utterance lengths; they did not, $t(94)=-1.05$, $SE=.042$, $p=.30$.

Languages.

Words were randomly assigned to concepts. I called this assignment of meanings to nonsense words “Language A.” One- and two-syllable words counterbalanced between nouns and verbs. After this “Language B” was then formed by randomly re-assigning noun nonsense word labels given in Language A to verb targets, and verb nonsense word labels given in Language A to noun targets. Languages A and B were counterbalanced between subjects.

Syntax.

Words in phrases were ordered in a noun-verb typology (as English uses), such as “[A] doctor [is] smoking.” Thus although words were initially presented in either a noun-verb or a verb-noun order, contextual utterances were always uttered in noun-verb order.

Images.

Ninety-six (48 noun and 48 verb) black-and-white line drawings images of various everyday items and actions, illustrated in referential isolation, were mostly found on the Internet.[21] These made up a convenience sample based mainly on two criteria: they were concrete nouns and observable action verbs, and a decent image of each could be obtained expediently. All objects and actions were of a basic semantic level, not too semantically specific but also not too general. Nouns were mostly animals (e.g., kangaroo) and professions (e.g., doctor), and a few inanimate objects (e.g., refrigerator); verbs were common, familiar actions that could be performed with parts of the human body, such as “to eat” (one exception was “to hatch”). Appendix A lists all the targets of the present study: 48 nouns, 48 verbs, and the 48 noun-verb pairs created with them.

[21] One major source of the images was an online database offering free line drawings of hundreds of objects and actions, along with naming norms, for language researchers by the Center for Research in Language at the University of California, San Diego: http://crl.ucsd.edu/~aszekely/ipnp/. Another resource was simply surfing the Internet using Google’s “images” option and filtering to search only black-and-white line drawings. Still other images were hand-drawn by two artistic research assistants: Kay Lee and Goldie Salimkhan.

Verb images may be considered images of the present participle. Importantly, verbs were all intransitive verbs, meaning they could be used without specification of a direct object (some could be considered both transitive and intransitive in nature, such as the verb *to write*). While transitive verbs require acting and acted-upon noun arguments, intransitive verbs only require an actor (noun argument in the subject position). By only using intransitive verbs, I could properly present verbs in two-word phrases that made sense without the need of additional information: each phrase described an actor performing an action.

Noun and verb "isolate" images (illustrated in referential isolation) conveyed just one concept per image, either a noun or a verb concept. An additional 48 "context" images contained two elemental concepts per image, always an actor performing an action. Some context images and verb isolate images necessarily contained a patient (receiver of action) as well as an actor, in spite of attempts to select only intransitive verbs for study. The noun and verb concepts in context images were the same as those in the isolate images. Thus, for example, one isolate image depicted surfing, and another depicted a computer. One of the context images depicted a computer that was surfing.

Revision of images was done in an effort to maximize name agreement. Toward this end, Adobe Photoshop and Windows Paint were used to delete background details, crop out any distracting or unnecessary details so that greater attention would be drawn to relevant parts, and add movement marks and lines of motion to verb images to suggest movement interpretations and make these images "come alive," as Figure 2 illustrates.

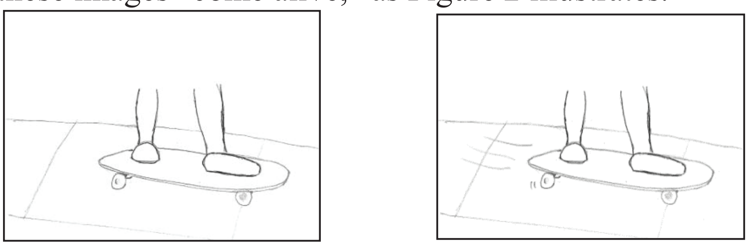

Figure 2. "Skateboarding" with (right) and without (left) movement marks. With movement marks the action component becomes more obvious and dramatic. This technique was used to make verbs in images more easily identified. To avoid copyright infringement, I present images which I drew rather than the actual skateboarding image used for experimentation.

Image heights and widths were measured and summed (height + width) for each image, and these sums were submitted for testing to ensure size differences between noun and verb images were not a confounding factor. Noun image dimension sums (877 pixels, SD=100) did not differ from those of verb images (864 pixels, SD=135), independent $t(94)=.516$, $SE=24.33$, $p=.607$.

Learning.

The learning program.

A Toshiba laptop computer (screen size: 19 inches diagonally) was used to present words, sounds, and images using Superlab 4.0. Details of how this program presented stimuli to participants follow.

*Events***.** A learning event was composed of an image and auditory stimulus (nonsense word) presented at the same time. The nonsense word's onset was purposely recorded with about 100 milliseconds of silence at the beginning of each sound clip so that words were not sounded simultaneously with the onsets of images. Events advanced over time at a rate of one event every three seconds.

Trials. Each learning trial was composed of a set of events, either two or three, presented sequentially, so each trial lasted either 6 or 9 seconds. Trials were presented consecutively as a continuous progression of images throughout each segment.

Target occurrences within trials. Participants learned words in ostensive segments and inferential segments within subjects. Among ostensive segments, targets were presented either once or twice in each learning trial; this variable, number of target occurrences per trial, was manipulated between subjects. Among participants who viewed targets with two occurrences per trial, each target's second occurrence marked the trial's end and thus would have allowed participants to parse trials and group words according to trials as pairs. However among participants who saw targets presented only once per trial, the presentation format was such that image pairs may not have been identified by participants as pairs, per se. That is, where no target occurred more than once per trial, trial end-points were less apparent, and participants presumably did not pass trials very well, which would have made grouping targets as pairs unlikely. This design cost was outweighed by its design advantage—enhanced control and comparability between methods of learning, and between levels of target occurrences. One-occurrence trials and inferential trials both contained two events; number of events controlled, so any recognition difference would owe to an effect of method of learning. One-occurrence and two-occurrence ostensive trials both contained ostensive presentations of the same isolate images; any difference to be found between these conditions would owe to the presence or absence of a context image in trials because the same isolate images (and characteristics of those images) were seen between these conditions.

Blocks. A block was composed of 8 trials. The same block was shown 6 times repeatedly with no breaks between block repetitions in each segment. Each time a block was shown, its trials occurred in a different randomized order.

Segments. A segment was composed of a block repeated six times. Each segment presentation lasted from five to seven minutes, and was entirely made up of either ostensive trials or inferential trials, but never both. Each segment was preceded by two practice trials demonstrating the pattern of that segment's image progression. Between segments, participants engaged in a distractor task, attempting a Sudoku puzzle, for 30 seconds to reduce learning fatigue, attention fatigue, and pro-, and retroactive interference from other learning blocks.

Two inferential learning segments and one ostensive learning segment were viewed at each learning session. The reason for twice the number of inferential to ostensive learning segments was to obtain equal numbers of data points from the two methods of learning. Twice as much ostensive data were collected per ostensive trial as inferential data were collected per inferential trial because each ostensive trial presented two words which could be tested, while each inferential trial only allowed inference of one word, so only that one could be tested from a given inferential learning trial. The two inferential learning segments were always presented consecutively to reduce number of instructional differences between segments. The ordering of segments was ostensive, inferential, inferential at one of two learning sessions, and inferential, inferential, ostensive at the other. Segment orders and order of segment orders were both counterbalanced between learning sessions.

Lists. Ninety-six words were evenly divided into two lists. The reason for two lists instead of one was to enable manipulation of the test-delay variable within subjects, and to reduce the number of targets per lists to make each learning session more manageable for participants. Each list was divided into three learning segments. Every participant learned both lists and thus saw six segments.

Learning schedules.

Participation occurred in two parts, both involving learning and testing. The testing effect (improved recall for already-tested items) was not an issue because no item was tested twice. Among 56 participants, half of the words were tested after a five minute delay, and the other half of the words were tested after one week. For the remaining 28 participants, the test delay was fixed at five minutes. To limit proactive and output interference for either list, a learning schedule was utilized in which 28 of the participants learned the first list at their first of two appointments and the second list at their second appointment one week later. The remaining 56 participants learned both lists within a single appointment. The two learning schedules are illustrated in Table 4. Schedule 1 might lead to more output interference due to learning both lists nearer in time, but less proactive interference due to the length of time intervening between learning both lists, while Schedule 2 might lead to more proactive interference but little output interference. Use of both schedules helped to even these effects out.

Table 4

Learning Schedules

Schedule	Learn	Delay [a]	Test	Learn	Delay	Test
1	list 1	1 week	list 1	list 2	5 minutes	list 2
2	list 1	5 minutes	list 1	list 2	1 week[b]	list 2

[a]The bold, zigzag line segment demarcates what occurred during the first (to the left of the line segments) and second appointments (to the right of the line segments), separated in time by one week.
[b]Half of the 56 individuals assigned to this learning schedule actually experienced a second five minute delay (not a one week delay as the diagram shows).

Instructions.

Participants read segment-specific instructions relating to each segment's learning condition. In the ostensive learning block the instructions read, "In this section, slides are ordered into TRIPLETS presented back-to-back: 1st – word 1 is spoken (you will see an illustration of it), 2nd – word 2 is spoken (and illustration), 3rd – a phrase is spoken containing those words again (and illustration). Both words are equally important." Among the sample of participants who were presented with ostensive conditions containing only one target occurrence per trial, directions read "In this section, slides are presented back-to-back." For the inferential learning blocks, instructions were as follows: "In this section, slides are ordered into PAIRS presented back-to-back: 1st – a word is spoken (you will see an illustration of it), 2nd – a phrase is spoken containing that word AND another word (and an illustration of them). Both words are equally important." After the above sets of instructions tailored to conditions were presented, some general instructions followed: "You do not need to respond. Just learn what the words mean. Later, you will be tested! Practice 2 triplets [inferential condition: "pairs"] first. The experimenter will guide you during this practice." After two training trials were shown, the following text appeared on the screen: "Can you tell the experimenter in your own words what you will be doing in this block?" Feedback was provided to clarify the instructions as necessary.

Testing.

The test was conducted using Superlab 4.0. English words, arranged vertically as numbered options, were presented in the center of the screen in Times New Roman 18-point font. At each item onset, a sound file presented a nonsense word from the most recent learning session. Two tests were given to assess learning of the two word lists. There were 96 test items testing all learned words.

Foils were chosen from among the learned stimuli so that all choices would be equally familiar. One of the foils was always a meaning that co-occurred with the target in its context image during learning. Each English word was offered four times at test, once as target, and three other times among foil options.

The participant's task was to indicate which English word was referred to by the spoken word (a forced choice paradigm) using four number keys ([1], [2], [3], and [4]) to designate choices. Two practice test items were given, the choices and targets derived from the learning practice trials. No breaks were given during testing. Participants' responses were scored by the presentation software as correct or incorrect.

Procedure

Participants were tested in a departmental lab space by any of seven research assistants or me. Participants who arrived at the research site in a timely manner were randomly assigned to a learning schedule. Those who arrived more than ten minutes late *and* indicated they had another engagement at the end of the hour were assigned to the schedule that would allow them to complete the first part of their participation within that hour (fewer than 10% of the sample). The effect of learning schedule on word recognition was assessed.

Upon entering the testing room, participants completed a consent form and filled in a language and biographical data form that asked for their sex, age, primary language, proficiency in English, other languages spoken, and proficiencies in those languages. Next participants were seated at a computer and the experiment was started. Participants completed two learning portions and two testing portions of the experiment according to the schedules shown in Table 4. When participants experienced the five-minute delay, they were told to play Tetris for five minutes before being tested. When they experienced the one-week delay, they left for that day and returned seven (minimum six, maximum eight) days later. Afterwards participants were debriefed and credited.

Results and Discussion

Analysis strategy

Three of the variables I manipulated in this experiment are the primary focus of this paper; I refer to these as my primary variables. The remaining variables (mostly non-manipulated) I refer to as secondary variables; these were measured foremost to ensure they were not confounded with my primary variables so as to allow safer interpretations of causes and effects. Second, any secondary variables that account for outcomes could add or subtract support for the RC hypothesis, which is important for theory development. Finally as some of these variables have not yet been shown to predict learnability, finding effects of them could yield novel contributions to the vocabulary acquisition literature. Given these variables were secondary to the purposes of this investigation, however, my analysis of them remains in the form of single variable assessments, or in some cases model components for providing greater assurance of interpretations of my primary variables.

Primary and secondary variables were initially assessed as single variables, but in order to provide a more compact and cohesive view of the effects of these study variables, I joined primary variables (method of learning, word class, delay, and all of their interactions) together into a single model to measure their effects relative to and in context with one another. Secondary "extraneous" variables were added one at a time to this initial model; if significant, variables joined and thus developed the model, but if insignificant ($p>.05$), variables were removed from the model before proceeding to subsequent model component testing, in the order displayed in Table 5. Model development was done to satisfy questions about whether primary variables' effects might still exist with potentially confounding variables controlled. Secondary variables were removed from model testing before proceeding to each subsequent test in order to keep all models as small and within a rule-of-thumb limit at all times. [22] Only variables that were by themselves significant predictors were tested in the primary factor model. This led to the development of a more developed model which included both primary and secondary variables to provide a fair estimate of the reliabilities, sizes, and directions of the effects of primary and secondary variables without either confounding interpretations of the others. Interpretations are then discussed.

I collected recognition responses for all words presented to all participants. Due to experimental programming error, certain test items were presented with options in which the target was absent. This occurred on 4 items under Language A, and on 10 items under Language B for two-thirds of the participants (these errors on the test were corrected prior to running the final one third of participants); data for these problematic test trials was removed before analyses. An alpha criterion of .05 was used to determine significance for all assessments. Logistic regression allowed me to take advantage of known qualitative and quantitative differences amongst stimuli, participants, and conditions, 23 factors in total, to test their predictive value. I used an alpha criterion of .05 to make decisions of significance. The initial survey of these 23 factors was exploratory by nature; determining significance while correcting for the error rate based on 23 tests requires differences to be highly reliable, but the power for such reliability was short. Therefore I present the results of these tests as a reference for future researchers interested in furthering work on these variables, but not as a final word on their effects. I also used an alpha criterion of .05 for reliabilities of model components to decide whether to retain or discard factors from models during model development.

Results of testing each factor in its own model are reported in Table 5 with Wald χ^2 values and p values to indicate reliability of improvement in individual model predictions over null hypothesis predictions based only on the grand mean. Effect sizes are only reported in developed multi-factor models because simple one-factor models tend to over- or under-estimate effect sizes when multiple effects are involved. Also shown in Table 5 are the interactions I tested between some of these factors and delay. For each test of an interaction, the two factors and their interaction factor were modeled (three factors) together.

[22] Just how many factors may be included in a model? A fairly common rule of thumb regarding sample sizes needed for regression analysis is $N > 10k$ when there are k predictors; this would allow up to 8 factors in my study with 84 participants. Green (1991) more conservatively proposed a rule of thumb where $N > 50 + 8k$ (this would translate to only 4 factors in my study); and Hosmer and Lemeshow (2000) advised $N = 20k$ in logistic regression (also translating to 4 factors in my study). But Vittinghoff and McCulloch (2006) found that such rules of thumb were too restrictive. In the present case, the sample size was 84, so in light of Vittinghoff and McCulloch's research, and given the fairly high number of observations per participant, I felt comfortable limiting the number of factors in models to eight or fewer. As the N : k ratio gets smaller, generalizations beyond the sample become riskier (Berger, 2003).

Only the p-values of the interaction components of these models are cited in the table.

Table 5
All 23 Factors Tested with Individual Models

	source	Main effects Wald χ^2	p	Means (SD), or % of cases	Interaction w/ delay p
Participant Factors	Age	0.01	0.94	20.22 (1.76)	n.t.[a]
	Sex	1.47	0.22	72.6% female	0.89
	Eng. as 1st lang	0.19	0.66	79.8% English 1st	0.05
	Eng. proficiency	0.12	0.73	9.64 (1.01)	n.t.
	Lang prof. sum	0.05	0.82	17.95 (4.42)	0.66
Condition Factors	Delay	306.68	0	67.1% 5-min	n.t.
	Class	0.2	0.66	50.2% nouns	0.007
	Method of learning	27.48	0	67.1% ostensive	0.01
	Order	5.18	0.02	50.0% n-v order	0.73
	Experiment lang	0.02	0.88	50.2% Language A	<.001
	Lists learned prior	63.32	1	65.7% no prior list	<.001
	Learning schedule	4.41	0.04	33.3% schedule 1	0.09
	Occurrences/trial	4.46	0.03	44.4% 1 occur/trial	0.57
Item Factor *Conceptual*	Category represent.	0.01	0.92	0.19 (.09)	0.24
	Familiarity	5.53	0.02	6.47 (.51)	0.63
	Imageability	7.56	0.01	6.68 (.40)	0.35
	Frequency	4.08	0.04	3.65 (.99)	0.93
Visual	Goodness depiction	1.89	0.17	4.85 (.26)	0.91
	Alternatives	7.36	0.01	4.65 (2.76)	0.63
	Strangeness	0	1	4.52 (1.87)	n.t.
	Name Agree-Isolate	2.85	0.09	0.88 (.14)	0.14
	Name Agree-Context	3.34	0.07	0.84 (.19)	0.86
Auditory	Utterance length	0.62	0.43	0.93 (.20)	0.06

[a] "n.t." refers to variables that were not tested either because they were unlikely to be significant, were correlated and thus partially accounted in another tested variable, or because they were not manipulated orthogonally with delay.

Initial model

The effects of word class, method of learning, delay, and their interaction terms—class x method, method x delay, class x delay, and class x method x delay—were initially tested in a 7 factor model. While the model was significant (p<.001), the three-way interaction was not. With the removal of the three-way interaction from the model, all six factors were reliably predictive (p<.05), as was the model itself, Wald $\chi^2(6)=348.65$, p<.001. This initial model is provided in Table 6, and its effects are illustrated in Figure 3. Effect sizes are measured here as odds ratios, a metric often used in logistic analyses. Odds ratios are relative to the reference value which is the minimum possible value on each variable's scale of measure. Appendix B offers a short introduction to odds ratios for readers unfamiliar with this metric of analysis.

Table 6

Initial Model Including Three Manipulated Predictors and Three Interactions

			Measured values	
Predictor	p	odds ratio	Min	Max
Word Class	<.001	0.83	noun	Verb
Meth of Learn	.006	0.62	ostensive	Inferential
Delay	<.001	0.28	5 minutes	1 week
Class x MOL	<.001	1.26	-	-
MOL x Delay	<.001	1.33	-	-
Class x Delay	<.001	1.32	-	-

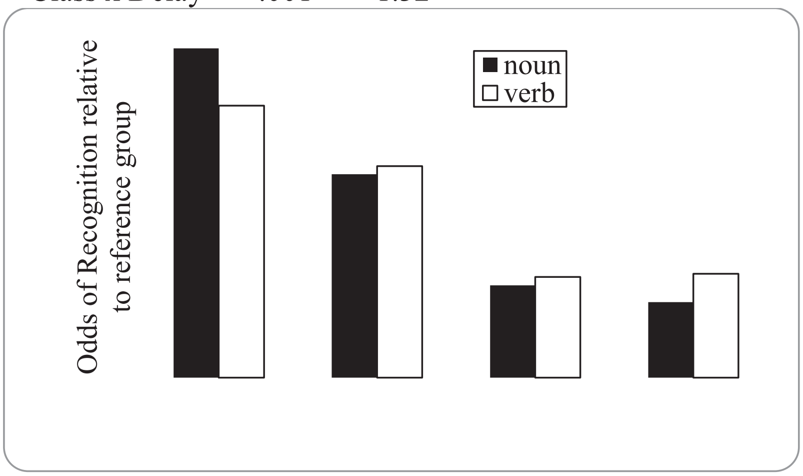

Figure 3. The effects of delay, method of learning, and word class on the odds of recognition according to the initial model. All effects are significant at p<.05. Odds values are relative to the reference condition (nouns learned under the ostensive condition and tested at five minutes) which has been arbitrarily set to 1.

Further model development

A total of seven factors were applied toward further development of the initial model (shown in Table 6). The first three factors were condition factors, and the last four were item factors. First I entered word order in the model.

Word order was nested at (only manipulated within) the ostensive level; therefore testing word order was done on only a subsample of the data. Not all variables could be included in this test because word order was not crossed with word class (so word class and its interaction terms were not tested). Therefore in order to proceed to testing other factors in the initial model beyond this test, I would have to remove word order from the model. However for the sake of testing word order, I temporarily removed word class from the model, as well as both of its interaction terms. The effect of order in this four-factor test was significant, $p=.02$ and indicated the verb-noun presentation order led to 1.17 times greater odds of recognition relative to the noun-verb presentation order . Its inclusion in the model did not greatly change estimated effect sizes (ORs) or reliabilities, and did not affect the directions of the effects of other factors in the model. Before proceeding to model tests with other variables, I removed word order and returned word class and its interaction terms to the model.

Next I added learning schedule, but it was not reliable, $p=.55$. Therefore I replaced learning schedule with number of occurrences per trial. This variable was significant in the model, $p=.006$, and its inclusion affected the model by increasing the effect size (but not direction) of method of learning, and reducing the reliability of the interaction between method of learning and delay, causing it to lose significance ($p=.07$). Therefore I replaced the insignificant interaction term with the significant number of occurrences term.

After this I tested four item factors in this revised model. Order of testing was somewhat arbitrary; I first added familiarity to the model and tested it. Familiarity was significant, $p=.02$; it did not affect the size, direction, or reliability of the other model factors. I thus retained familiarity in proceeding to test the next item factor, imageability. Imageability came close but did not quite meet my significance criterion, $p=.06$; at the same time imageability reduced the significance of familiarity to $p=.54$. This was not a surprise; past research has shown imageability and familiarity, as well as frequency, are highly correlated (Bird, Franklin, & Howard, 2001), and these terms were also highly correlated in the present data, all $r>.6$, all $p<.001$. As imageability was nearly significant even when controlling for familiarity, I retained imageability, and replaced familiarity with frequency, the next factor I tested. Frequency was not significant, $p=.88$, so I removed it from the model. Finally I tested number of alternative interpretations of images in the model. Number of alternatives was significant, $p=.03$, did not change the directions of other effects, and did not greatly change their reliabilities or effect sizes. Therefore this variable was accepted into this more fully developed model of word learning for this study. The final model given in Table 7 follows the rule of thumb $N > 10k$ (8 model factors, including interaction terms), and was highly significant ($p<.0001$). The same factor effects illustrated earlier for the initial model are again illustrated in Figure 4 but with the effects of number of target occurrences per trial, word imageability, and number of alternative interpretations controlled. These modeled effects are better estimated than the prior estimations (the initial model); the final model's effects are estimated apart from the effects of number of occurrences, word imageability, and number of alternative interpretations.

Table 7
Final Model with Eight Factors (4 Manipulated, 2 Interactions, and 2 Non-manipulated)

			Measured values	
Predictor	p	odds ratio	Min	Max
Word Class	.04	0.86	noun	verb
Meth of Learn	<.001	0.51	ostensive	inferential
Delay	<.001	0.31	5 minutes	1 week
Class x MOL	.04	1.24	-	-
Class x Delay	.006	1.33	-	-
Occurrences	<.001	0.71	1	2
Word Imageability	.01	1.17	4.43[a]	7
No. of Alternatives	.03	0.98	1[b]	17

[a] The minimum value possible on this scale was 1.
[b] the minimum value possible on this scale was 0.

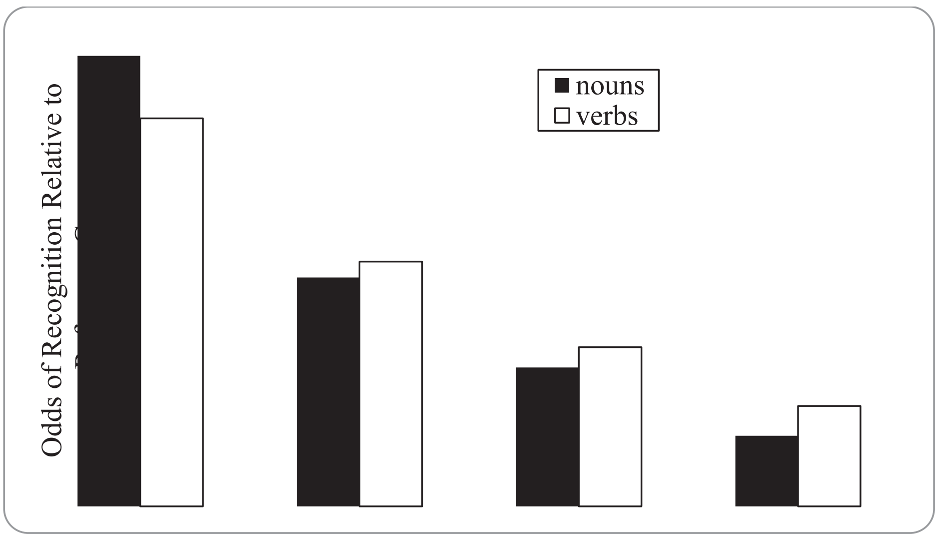

Figure 4. The effects on the odds of recognition by word class, method of learning, delay, and the word class interactions with method of learning and with delay, based on the final model. Odds of recognition are relative to the reference

group, nouns learned under ostensive conditions and tested at five minutes, where the odds have been arbitrarily set to 1.

Main and simple effects based on the developed model.

Delay. There was an effect of delay. With logistic regression, main effects are not usually interpreted when their factors are involved in interactions. The effect of delay was obviously detrimental to recognition, but delay hurt recognition of nouns more than of verbs. Nouns stood .31 (about a third) fewer odds of identification at one week compared to five minutes. However verbs stood only .41 fewer odds of identification at one week compared to five minutes.

Method of learning. Method of learning interacted with word class such that the advantage of ostensive learning over inferential learning, observable from Figure 4, was not evenly distributed between nouns and verbs. Among verbs the odds of recognition under inferential learning conditions were only .63 times as great as they were under ostensive learning conditions. However among nouns, inferential learning led to .51 times worse odds of recognition compared to ostensive learning (i.e., ostensive learning improved recognition odds by a factor of about 2).

Perhaps inferential learning is better suited to verbs than nouns. Verbs are relational concepts (verbs relate objects and a change of state to one another), and inferential learning highlights relationships between image pairs (the repeated and novel element in the context image is seen only in relation to the isolate image presented just prior). Perhaps inferential conditions prime relational thinking, thereby highlighting verbs more than nouns. However another hypothesis even seems more probable to me, and fits well with the RC hypothesis: perhaps inferential learning of verbs led to greater recognition because inferential learning of verbs was supported by prior ostensive presentation of nouns, leading to greater context clarity for verb inferences. Inferential learning of nouns was supported by prior ostensive presentation of verbs, but prior ostensive presentations of verbs may not have been very helpful; prior verb presentations may have even added ambiguity to context clarity for noun inferences.

Word class. The noun bias was observed under a single condition only—when learned under ostensive conditions *and* when tested at five minutes, the noun bias appeared such that nouns had 1.16 greater recognition odds compared to verbs. When tested at one week, this advantage was reversed such that verbs stood 1.15 greater recognition odds than nouns. When tested at five minutes under inferential conditions, verbs stood 1.07 greater recognition odds than nouns. And when compounding these verb-friendly conditions, the verb bias was at its strongest—verbs stood 1.42 times greater recognition odds than nouns when inferentially learned and recognized at one week.

Number of occurrences. The number of times a target occurred within each trial was predictive of recognition success, but in the opposite direction to that expected. Two occurrences per trial (with one isolate and one context image) led to .71 times lower recognition odds compared to only one occurrence per trial (with one isolate image). How could this have happened? My colleagues have suggested this may have occurred because isolate images of verbs may have been seen as actor- verb combinations rather than purely as verbs. (Most isolate verb images were necessarily illustrated with an actor.) Hence presenting a verb learned in one actor-context again in another actor-context might have led to confusion about the meaning of that word (i.e., Does "jev" mean boy-dribbling or alligator-dribbling?) This theory is essentially the same as Tomasello's verb-island hypothesis (year)—the theory that young children's initial understanding of verb meaning is under-extended to include only a very narrow set of actors (perhaps just one) or patients (perhaps just one) or tools (perhaps just one). In other words, the immature form of "verb knowledge" is highly specific to a particular noun context—so much that a new verb label may be considered an "actor-verb" by the learner, initially. Appropriate generalization of a verb's meaning to the appropriate number of actors, patients, and tools may only occur gradually, after extended exposure to varied verb circumstances. Hence my colleagues and I take this unusual effect of the number of occurrences as evidence of an adult form of the verb-island experience, a starting point from which vocabulary meanings can mature. Hypothetically, a learner could begin to identify a verb's exact meaning by over-generalizing and work backwards to appropriate specificity. Child language development studies reveal children start specific and grow general; adults may show the same pattern, but more work is needed to support this interpretation.

Word imageability. Word imageability had a positive effect on the odds of recognition, elevating the odds by a factor of 1.17 for every one unit increase in imageability from 4.43 (the lowest observed word imageability rating on the scale of 1 – 7) to 7.00. This means the easier it is to visualize a word's meaning in one's native language, the easier it is to learn in a foreign language.

Number of alternatives. The effect of the number of alternative interpretations on recognition was as expected: images for which people offered more alternative interpretations were associated with lowered probabilities of recognition. This measure of image quality was more reliably predictive of learning outcomes than were rated goodness of depiction ratings and name agreement measurements. This effect adds support to the idea that learning is largely a matter of how clearly meanings are presented to learners (i.e., the RC hypothesis).

Other findings (outside of the developed model)

Some variables appeared significant in their own models but failed significance testing when modeled with the primary factors word class, method of learning, and delay. Frequency did not meet significance when primary factors were accounted for. Familiarity met the significance criterion until imageability was added, which accounted for the predictive effect of familiarity. These therefore were not good predictors of word learning. The effect of order was significant in its own model as well as when modeled with word class and delay (i.e., with their effects controlled). It was excluded from the developed model to avoid model complexity, as this variable was nested within only one of the methods of learning.

I did not attempt to include interactions between delay and secondary variables because I did not want to jeopardize the integrity of models by overloading them with factors and exceeding the rule of thumb (N=10k) for the number of factors modeled, given N=84. However future experimenters are urged to better understand these observed interactions. The interaction between experiment language (Language A vs. B) and delay revealed that at five minutes there was a recognition advantage for learners of Language A (OR=.78), but at one week there was a recognition advantage for learners of Language B (OR=1.14). Although these differences were not significant at either delay taken individually, the reversal of this difference from one delay to the other was very reliable. For the sake of curiosity I tested this interaction holding word class and method of learning constant; its effect remained highly reliable ($p<.001$). Perhaps this effect stems from temporally dynamic specific item effects (word learning difficulty as a product of the way they sound in combination with the meanings they represent). That is, perhaps words that seem easy to remember really are easier to remember at five minutes, but are forgotten more quickly, too.

Following in this same pattern, the interaction between delay and the number of lists learned prior to a given word's list was highly reliable ($p<.001$), even when holding word class and method of learning constant. This interaction showed that at five minutes, there was a slight difference between learning words in a first or second list (OR=1.08) indicative of better learning of list 2 words; at one week, however, list 2 words were at a disadvantage (OR=.85). Although these effects were non-significant at either delay taken individually, the reversal in the pattern of differences was highly reliable. Perhaps proactive interference caused faster forgetting of words learned in list 2, but its effect on the forgetting rate may not have been strong enough to see after only five minutes. Also there may not have been sufficient power to find an interference effect at one week (fewer participants, only 56, were tested at one week).

Finally, again following the same interaction pattern as found between delay
and experiment language and between delay and number of lists learned prior, the interaction between learning schedule and delay was highly reliable ($p<.001$), even when controlling for word class and method of learning. This interaction revealed learning schedule 1 (learning sessions separated by a week) led to better performance when tested at five minutes compared with learning schedule 2, but learning schedule 2 (learning sessions separated by five minutes) led to better performance at one week compared with learning schedule 1. This pattern of effects violates expectations based on proactive and retroactive interference effects. Learning schedule 2 should have caused greater proactive interference on words tested at one week, and similarly learning schedule 1 was expected to cause a very small amount of proactive interference for words tested at five minutes. One explanation for this pattern of findings is a beneficial practice effect. In this view, learning schedule 1 led to better performance than learning schedule 2 at five minutes because under learning schedule 1 the five minute test was always on the second list, and was always the second test taken. Learning schedule 2 led to better performance than learning schedule 1 at one week because under learning schedule 2 the one week test was always over the second list learned, and was therefore the second test taken. Thus perhaps participants improved their performance on the second list learned because they learned how to learn words for the type of test given (multiple choice English word targets).

General Discussion

The model developed from this experimental data shows that learning can be described as a product of learning conditions, the target words, and the learning materials. Studies of young learners in most language across the globe have demonstrated a noun bias dominates early repertoires (e.g., Gentner, 1982). My aim in this study was to test whether adults show indication of a noun bias, too. The results of this study, taken together, do not support a generalized noun bias in adulthood; thus I did not find support for a noun bias as others have observed among young language learners. However this experimental paradigm is not exactly a mirror image of the learning challenges undergone by early language learners. Young learners must not only learn which labels map onto which referents, but they must also sift for content words through non-referential words such as closed-class words; learn or not mind that some words are reflexive; learn their language's syntax; and perhaps most importantly, they must identify which aspects of a state change is referenced by a verb label, and which aspect of an object or group is referenced by a noun label. These last challenges mark an important difference between early word learners and mature second language learners: for the most part, second language vocabulary can be supported by already-acquired first language vocabulary (arguably most words have equivalents or near-equivalents between any pair of languages), but first language learners do not have this learning support. Thus it may not be that surprising to some that the noun bias, virtually universal among children acquiring language for the first time, may not generalize to adults learning language a second time.

Beyond arguing that adults do not experience a noun bias in second language learning of the same kind or magnitude as children, I shall also propose that in vocabulary development, recognition likelihood is a dynamic construct that varies over time. This was most clearly seen in the noun advantage under ostensive conditions at five minutes; under the same learning conditions, this became a verb advantage at one week. One hypothesis formed post hoc for the temporal dynamics of word learning is this: that which makes learning easy in the short run hurts learning in the long run. Although referential clarity is argued to be a key to successful learning, greater referential clarity might also permit faster, shorter processing time which may set that learned material on a faster forgetting course. In this proposal, shorter processing time contributes to better performance in the short run, but longer processing time contributes to better performance in the long run.

A competing but non-endorsed hypothesis is that the learning dynamics observed in this study were the result of floor effects at one week. One reason I do not agree with this latter hypothesis is the observed interaction between experiment language and delay which indicated that at five minutes, performance by those who learned Language A seemed greater but at one week performance by those who learned Language B seemed better. That is, the direction of the experiment language effect reversed order. I interpreted this interaction in terms of more numerous positive specific item effects in Language A causing greater performance at five minutes by those who learned Language A due to faster processing which led to greater recognition likelihood at short delays, but also faster forgetting, which could account for the inferior performance of those assigned to Language A versus Language B at one week.

Caveats

Due to counterbalancing mistakes discovered only after all data were collected, one-third of the 96 words were only learned inferentially. The other two-thirds were properly counterbalanced between methods of learning between participants. Therefore any advantage or disadvantage of inferential learning could either be due to the method of learning or due to a greater number of easy or difficult words learned inferentially only. However the models I developed helped to annihilate such a critique by showing that any variables that were by themselves contributors to learnability were, in the developed model, not confounded with other modeled factors, by virtue of their mutual inclusion in the model. Another problem with this experiment that I only recognized after data collection was that order of words in ostensive trials was not counterbalanced by target; each target word was presented in the context of either the noun-verb order or the verb-noun order. Therefore the effect of word order could have been confounded by chance assignment of more learnable words to one of the two word orders, and therefore its effect, reported above, should be taken with a grain of salt.

As mentioned earlier, an antithetical view of learning with temporal dynamics is that these temporal dynamics are only a product of floor effects of measurement at one week. Elevating performance at the later delay might improve sensitivity to these effects. This might be done by testing after fewer days, by teaching fewer targets, or by teaching with more, unique isolate and context images rather than the same ones repeated many times. Future researchers are encouraged to use one or more of these techniques to elevate learning at later delays to provide greater measurement sensitivity and rule out this antithetical view.

Future directions

The view that there are temporal dynamics at play in measuring vocabulary recognition at different delays is worth investigating further. Future researchers might manipulate processing time to see how it affects performance after shorter and longer delays to test this temporally dynamic view of vocabulary recognition. If recognition measurement is really temporally dynamic, an implication for word learning may be to slow processing time, or more likely deepen it (not just slow it), by requiring learners to infer word meanings within or across learning situations, or by some other creative method.

Another avenue of research occurred to me whilst making numerous image revisions to try to make verbs as identifiable as nouns (but the verbs were still less accurately named than the nouns). Over the course of these revisions, my research team and I learned that verbs usually are not easily recognized in images unless they are "animated" with certain, often-used tricks employed by cartoonists and artists. We used graphic motion cues including lines to indicate from whence movements originated, and marks near moving parts, to introduce dynamism into still images. While these symbols are not found in the ecology of the real world, I perceived they facilitated interpretation of motion from still images. I did not measure name agreement rates before and after verb image doctoring, so improved name agreement by employing these symbolic devices for now remains an impression. If these and other symbolic devices really did elevate name agreement, an educational application based on this possible effect and based on the RC hypothesis could be proposed. Conveying word meanings with images and without need to reference any other language could be very useful for instructors trying to create more immersive learning conditions. Also the use of images rather than words could allow materials to be presented with almost any population with few needed changes. And because images are highly memorable (Lutz & Lutz, 1978), image media are probably a highly useful means of teaching foreign vocabulary.

Word learning research with implications for adults learning second language is a fruitful area of investigation. Future researchers are encouraged to consider long term as well as short term benefits of various methods of vocabulary acquisition with an eye toward improving efficient learning practices inside and outside of language classrooms. Finally, whether learning by ostensive or inferential methods, how the number of examples affects learning is likely to be of great research value. My own future research plans involve manipulating number of learning examples orthogonally with word usefulness to see how these variables relate to one another and to word learning. To develop our understanding of how these and other variables contribute to vocabulary acquisition is paramount to acquiring better understanding of language learning and of the language learner.

APPENDIX A

Nouns		Verbs	
alligator	hippo	to bathe	to pray
angel	horse	to bungee jump	to read
apple	kangaroo	to clap	to rock climb
armadillo	king	to cook	to run
astronaut	mail carrier	to cry	to shout
baby	monkey	to dig	to sing
bear	moose	to dribble	to sit
bird	nurse	to drink	to skateboard
boy	octopus	to eat	to ski
car	penguin	to fish	to sled
cat	pig	to golf	to sleep
computer	police officer	to hatch	to smoke
cow	princess	to hug	to sneeze
deer	rabbit	to iron	to snort
doctor	refrigerator	to jump	to spin
dog	robber	to jump rope	to surf
dragon	sailor	to kayak	to swim
duckling	sheep	to kiss	to talk
elephant	spider	to knit	to type
fire fighter	strawberry	to laugh	to walk
fish	telephone	to mop	to wave
flower	turtle	to paint	to whisper
frog	witch	to parachute	to wink
hedgehog	zebra	to point	to write

Noun-verb Pairs			
alligator dribbling	cow digging	hippo knitting	princess ironing
angel hugging	deer winking	horse snorting	rabbit laughing
apple jump roping	doctor smoking	kangaroo skiing	refrigerator running
armadillo climbing	dog singing	king typing	robber shouting
astronaut bungee jumping	dragon hatching	mail carrier praying	sailor kayaking
baby drinking	duckling crying	monkey clapping	sheep golfing
bear bathing	elephant sitting	moose sledding	spider parachuting
bird reading	fireman pointing	nurse mopping	strawberry walking
boy skateboarding	fish kissing	octopus eating	telephone fishing
car talking	flower sneezing	officer writing	turtle spinning
cat sleeping	frog swimming	penguin painting	witch whispering
computer surfing	hedgehog waving	pig cooking	zebra jumping

APPENDIX B

Interpreting odds ratios for categorical variables requires deciding (sometimes arbitrarily) on a reference value for a given variable. For example the reference value of word order was noun-verb. With the reference value established, the odds ratio can be understood as the ratio of the odds of outcome when the predictor is at its alternate value (in this example, the verb-noun order) over the odds of outcome when the predictor is at its reference value. In the odds metric, odds ratios may be flipped to describe the odds of the reference value from the perspective of the alternate value. Using the word order effect found in the over-sized model of word learning for Study 2 in Table 3.4 as a concrete example, the effect size of word order is 1.17. This means that this model predicts when words are ordered verb-noun, the odds of their recognition is 1.17 times greater than when they are ordered at their reference value, noun-verb.

Interpreting odds ratios for continuous variables is only a little more complex. The reference value is always defaulted at the bottom of the scale. Using imageability as a concrete example, imageability values lay on a 1 – 7 scale, thus 1 was its reference value. The odds ratio of imageability was also 1.17 coincidentally, but on a scale of 1 – 7, the change in odds from minimum to maximum imageability values this is a larger change in odds (i.e., effect) than the effect of word order (whose odds were also 1.17). The odds ratio can be understood as the rate of change in predicted odds along a variable's scale. Thus with every one unit increase in imageability (say from 6.00 to 7.00) there is an associated change in the odds of successful recognition by a factor of 1.17. This odds ratio applies across the entire spectrum of measured (and unmeasured) values, and is a description of the effect size associated with a change in imageability of one incremental unit on the measured scale. That is, the odds of recognition at 7.00 are 1.17 times greater than the odds at 6.00, and the odds at 6.00 are 1.17 times greater than the odds at 5.00, etc. Improving imageability from 5.00 to 7.00, the model predicts, is associated with improvement from the odds of recognition at 5.00 (whatever that might be) by 1.17 x 1.17, or 1.37 times greater. Therefore although this odds ratio appears small, it is a hefty effect size when considering the improved odds of recognition along the entire spectrum of predictor values. In this example, the entire spectrum of imageability values was rather limited, but predicted recognition when imageability was at its maximum 7.00, compared to when it was at its minimum, 4.42 (so 7.00 is 2.58 units higher) is calculated as $1.17^{2.58}$, or 1.50 times greater predicted odds. Thus all other things being constant, "dog" (whose imageability was 7.00) was 1.5 times more likely to be recognized from its nonsense word cue than "hedgehog" (whose imageability was 4.42).

Understanding odds ratios with interactions is less straightforward than with main effects. When an interaction is significant, one should not interpret the involved main effect odds ratios by themselves because these values are displayed in output at their values when all other variables are held at their reference values (CRMportals, 2006), and they should be qualified this way. Word order and imageability could be deciphered simply because they were not part of any interaction factors. However the effect of method of learning was involved in at least one other interaction, so when describing the effect of method of learning, one must qualify this description by the level of the other variables it interacted with. Thus the effect of method of learning, as displayed in Table 3.4, was .49 for nouns (the reference value of the word class variable) only; for verbs the effect of method of learning was different, namely .49 x 1.32 (the interaction factor's odds ratio) = .65. In other words, the model predicts that when nouns are learned inferentially, they have .49 times lower odds of successful recognition than when they are learned ostensively, but for verbs, this negative effect of inferential learning is a little milder—verbs learned inferentially are only .65 times less likely to be recognized than when they are learned ostensively.

It is possible to convert odds ratios into likelihoods by the formula: (odds ratio / 1 + odds ratio) = likelihood. An odds of 1.00 means no effect, so the likelihood of success under this would be (1/(1+1)=.50) exactly 50% when all other model factors are controlled. Applying this to the effect of method of learning, the model specified that when all other variables are controlled, nouns were (.49 / 1 + .49 = .329) about 33% likely to be recognized when learned inferentially, and verbs were (.65/1+.65 = .394) about 39% likely to be recognized when learned inferentially. Calculating likelihoods of success for reference values from the perspective of alternate values involves flipping odds ratios. The likelihood of successful recognition of a noun learned ostensively is ((1/.49)/1+(1/.49) or 2.04/1+2.04 = .671) about 67%, and the likelihood of successful recognition of a verb learned ostensively is ((1/.65)/1+(1/.65) or 1.54/1+1.54 = .606) about 61%. Notice that averaging likelihoods of noun recognition at ostensive (67%) and inferential (33%) results in 50% average likelihood (i.e., no effect), and the same is true of verbs learned ostensively (61%) and inferentially (39%), which average 50%; this math indicates that model odds values are provided as values when all other values, including the model's intercept, are controlled so that they may be ignored.

References

Akhtar, N., Jipson, J., & Callanan, M. A. (2001). Learning words through overhearing. *Child Development, 72*(2), 416-430.

Appleman, I. B. & Mayzner, M. S. (1981). The letter-frequency effect and the generality of familiarity effects on perception. *Perception & Psychophysics, 30*(5), 436-446.

Au, T., Dapretto, M., & Song, Y. (1994). Input vs. constraint: Early word acquisition in Korean and English. *Journal of Memory and Language, 33*, 567-582.

Balota, D., Yap, M. J., & Cortese, M. J. (2006). Visual word recognition: The journey from features to meaning (A travel update). In M. Traxler, & M. Gernsbacher, Eds.), *Handbook of Psycholinguistics (2nd Ed.)* (Chapter 9). San Diego, CA: Academic Press.

Bird, H., Franklin, S., & Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavior Research Methods, Instruments, & Computers, 33*(1), 73-79.

Bjork, R.A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe and A. Shimamura (Eds.), *Metacognition: Knowing about knowing.* (pp.185-205). Cambridge, MA: MIT Press.

Brown, G. D & Watson, F. L. (1987). First in, first out: Word learning age and spoken word frequency as predictors of word familiarity and word naming latency. *Memory & Cognition, 15*(3), 208-216.

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development, 15,* 17–29.

Center for Research on Languages – International Picture-Naming Project. Image retrieved on June 2, 2012 from http://crl.ucsd.edu/experiments/ipnp/

CRMportals Inc. (2006). Interaction terms vs. interaction effects in logistic and probit regression. Retrieved on 1/17/2012 from http://www.crmportals.com/crmnews.

Ellis, A. W. & Morrison, C. M. (1998). Real age-of-acquisition effects in lexical retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(2), 515-523.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. *Technical report no. 257* No. BBN-R-4854

Gentner, D. (2006). Why verbs are hard to learn. In K. Hirsh-Pasek, & R. M. Golinkoff (Eds.), *Action meets word: How children learn verbs.* (pp. 544-564). New York, NY, US: Oxford University Press.

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition, 73*, 135-176.

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C (2006). Hard words. *Language learning and Development, 1*(1), 23-64.

Gopnik, A., & Choi, S. (1990). Do linguistic differences lead to cognitive differences? A cross-linguistic study of semantic and cognitive development. *First Language, 10*(3), 199-215.

Gopnik, A., & Choi, S. (1995). Names, relational words, and cognitive development in English and Korean speakers: Nouns are not always learned before verbs. In M. Tomasello, & W. E. Merriman (Eds.), *Beyond names for things: Young children's acquisition of verbs.* (pp. 63-80). Hillsdale, NJ: Lawrence Erlbaum Associate.

Greenfield, P. M. & Alvarez, M. G. (1980) Exploiting nonverbal context to promote the acquisition of word-referent relations in a second language. *Hispanic Journal of Behavioral Sciences, 2*(1), 43-50.

Greenfield, P. M. & Smith, J. (1976). *The structure of communication in early language development*. New York: Academic Press.

Horst, J. S. & Samuelson, L. K. (2008). Fast mapping but poor retention among 24-month-old infants. *Infancy, 13*(2), 128-157. doi: 10.1080/15250000701795598

Jaswal, V. K., & Markman, E. M. (2003). The relative strengths of indirect and direct word learning. *Developmental Psychology, 39*, 745-760.

Kauschke, C. & von Frankenberg, J. (2008). The differential influence of lexical parameters on naming latencies in German. A study on noun and verb image naming. *Journal of Psycholinguistic Research, 37*, 243-257.

Kersten, A. W., Smith, L. B., & Yoshida, H. (2006). Influences of object knowledge on the acquisition of verbs in English and Japanese. In K. Hirsh-Pasek, & R. M. Golinkoff (Eds.), *Action meets word: How children learn verbs.* (Chpt. 19). New York, NY, US: Oxford University Press.

Knight, S. (1994). Dictionary: The tool of last resort in foreign language reading? A new perspective. *Modern Language Journal, 78*, 285-299.

Krashen, S. & Scarcella, R. (1981). On routines and patterns in language acquisition and performance. *Language Learning, 28*(2), 284-300.

Lutz, K. A. & Lutz, R. J. (1978). Imagery-eliciting strategies: Review and implications of research. *Advances in Consumer Research, 5,* 611-620.

Markman, E. M. & Wachtel, G. F, (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology, 20(2),* 121-157. doi:10.1016/j.physletb.2003.10.071

Masterson, J. & Druks, J. (1998). Description of a set of 164 nouns and 102 verbs matched for printed word frequency, familiarity and age-of-acquisition. *Journal of Neurolinguistics, 11*(4), 331-354. doi:10.1016/S0911-6044(98)00023-2

Naigles, L. R. (1990). Children use syntax to learn verb meanings. *Journal of Child Language, 17*(2), 357-374.

Pinker, S. (1994). *The language instinct: How the mind creates language*. New York, NY: William Morrow & Company, Inc.

Sandhofer, C., & Smith, L. B. (2007). Learning adjectives in the real world: How learning nouns impedes learning adjectives. *Language Learning and Development, 3*(3), 233-267.

Shady, M., & Gerken, L. (1999). Grammatical and caregiver cues in early sentence comprehension. *Journal of Child Language, 26*(1), 163-175.

Snedeker, J. & Gleitman, L. R. (2004). Why is it hard to label our concepts? In D. G. Hall and S. R. Waxman (Eds.), Weaving a lexicon (pp. 257-293). Cambridge, MA: MIT Press.

Székely, A., D'Amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G.,& Bates, E. (2003). Timed picture naming: Extended norms and validation against previous studies. *Behavior Research Methods, Instruments, & Computers 35* (4), 621-663.

Tardif, T. (1996). Nouns are not always learned before verbs: Evidence from Mandarin speakers' early vocabularies. *Developmental Psychology, 32*(3), 492-504.

Tardif, T., Shatz, M., & Naigles, L. (1997).Caregiver speech and children's use of nouns versus verbs: A comparison of English, Italian, and Mandarin. *Journal of Child Language, 24,* 535-565.

Tomasello, M. (1992). The social bases of language acquisition. *Social Development, 1*(1), 67-87.

Verspoor, M. & Lowrie, W. (2003). Making sense of polysemous words. *Language Learning, 53*(3), 547–586.

Vitevitch, M. S. & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language, 40,* 374–408.

von Restorff, H. (1933). Über die wirkung von bereichsbildungen im spurenfeld. *Psychological research, 18*(1), 299-342.

Waxman, S. R., & Lidz, J. L. (2006). Early word learning. In D. Kuhn, R. S. Siegler, W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology: Vol 2, Cognition, Perception, and Language (6th ed.).* (pp. 299-335). Hoboken, NJ: John Wiley & Sons Inc.

Woodward, A. L., Markman, E. M. & Fitzsimmons, C. M. (1994). Rapid word learning in 13- and 18-month-olds. *Developmental Psychology*, *30*, 553-566.