Journal of Memory and Language 97 (2017) 1–16

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Phrase frequency effects in free recall: Evidence for redintegration

Cassandra L. Jacobs [a,*], Gary S. Dell [a], Colin Bannard [b]

[a] University of Illinois at Urbana-Champaign, United States
[b] University of Liverpool, United Kingdom

ARTICLE INFO

Article history:
Received 22 February 2017
Revision received 19 June 2017

Keywords:
Phrase frequency
Redintegration
Memory
Language production
Recall

ABSTRACT

Three experiments examined the effects of word and phrase frequency on free recall. Word frequency did not affect word recall, but when participants studied and recalled lists of compositional adjective-noun phrases (e.g. *alcoholic beverages*), phrase frequency had a consistently beneficial effect: both words from frequent phrases were more likely to be recalled than for infrequent phrases, providing evidence that long-term memory for phrases can aid in pattern completion, or redintegration. We explain these results and those of a previous study of phrase frequency effects in recognition memory (Jacobs, Dell, Benjamin, & Bannard , 2016) by assuming that the language processing system provides features that are linked to episodic contexts. Recall tasks map from these contexts to linguistic elements, and recognition maps from linguistic elements to contexts. Word and phrase frequency effects in both memory tasks emerge both within the language processing system and from multiple stored episodes, and the fact that the representations of phrases are tied to knowledge of their component words, rather than being representational islands.

Published by Elsevier Inc.

Introduction

In many linguistic tasks, phrase frequency effects mirror word frequency effects. Common words (e.g. woman) and phrases (e.g. alcoholic beverage) are easier to acquire, understand and produce than uncommon words and phrases (Arnon & Cohen Priva, 2013, 2014; Arnon & Snider, 2010; Bannard & Matthews, 2008; Bybee, 2006; Janssen & Barber, 2012; Morgan & Levy, 2016; Siyanova-Chanturia, Conklin, & Van Heuven, 2011). The existence of phrase frequency effects demonstrates that the language processing system pays attention to multiword linguistic units. Frequency effects for individual words have typically been accounted for by either positing a lexical entry that keeps track of something like the count of times a person has encountered a linguistic category, or individual memories (exemplars, episodes, or instances) for each of those experiences. Because phrases include a temporal or grammatical relationship between multiple words, it is less clear how phrases might be represented in long-term memory. The present study addresses this question.

One way to explain phrasal frequency effects and phrase representation in general is to propose the existence of a lexically-specific but usage-event-independent representation of the phrase, such as a "node" (e.g. MacKay, 1982) or "superlemma" (e.g. Sprenger, Levelt, & Kempen, 2006) that contains information about its category (e.g. noun phrase, for an adjective-noun combination) and connects to representations of its component words (e.g. Copestake et al., 2002). The frequency of a phrase could be stored with this lexical entry, or it could arise from the number of stored episodes that contain or point to it. Alternatively, phrases could lack explicit discrete representations entirely, in line with theories and computational models that encode all words and phrases implicitly in network weights (Baayen, Hendrix, & Ramscar, 2013; Baayen, Milin, Đurđević, Hendrix, & Marelli, 2010; Seidenberg & McClelland, 1989).

Some recent work has looked into whether phrase frequency effects arise from speakers of a language tracking the episodic representations of events in which a phrase is experienced. Jacobs, Dell, Benjamin, and Bannard (2016) tested whether people showed the same sensitivity to phrase frequency in recognition memory as they are known to have for words. In single-word recognition memory experiments, words that a participant has rarely experienced over the course of their life (low frequency words) have fewer episodic memories in long-term memory, and yet are more accurately discriminated from lures than high frequency words are (Glanzer & Adams, 1985; Hintzman, 1988; Reder et al., 2000). This paradoxical effect of word frequency can be explained by noting that to judge a test word as "old" in a recognition task, the participant may retrieve the episode in which the word was studied.

* Corresponding author at: 603 E. Daniel St., Department of Psychology, University of Illinois Urbana-Champaign, Champaign, IL 61820, United States.
E-mail address: cljacob2@illinois.edu (C.L. Jacobs).

http://dx.doi.org/10.1016/j.jml.2017.07.003
0749-596X/Published by Elsevier Inc.

When that word is low frequency, there are fewer other episodes of it to hinder the search for the crucial experimental episode. Jacobs et al. reasoned that, if adjective-noun phrases have their own episodic memories that contribute to memory in the same manner, then low frequency phrases like *psychic nephew* should also be more accurately recognized than high frequency phrases like *alcoholic beverages*. Surprisingly, they found that high and low frequency phrases were recognized equally well, but that recognition memory improved when the noun in a phrase was uncommon (e.g. *wizard* improves memory for *handsome wizard*). That is, the ability to discriminate new from old phrases, as reflected in a higher hit rate and a lower false alarm rate, was unaffected by phrase frequency, but it was benefited by low-frequency nouns within phrases. They concluded that recognition judgments for phrases are more influenced by the number of episodes containing particular words within the phrase, as opposed to the entire phrase. This is so because individual words are necessarily much more common than phrases. Thus, the many episodes sharing a word with a test phrase are more potent sources of interference in the recognition process than the few episodes containing the entire phrase.

This finding from Jacobs et al. (2016) provides evidence that phrasal processing is at least partially compositional, in that judgments about *psychic nephew* are influenced by memories of events of psychic things that are not nephews and nephews that are not psychic. However, the study also found that participants tended to say they had studied the more common phrases (e.g. *alcoholic beverages*), as evidenced by a bias to respond "yes" with increasing phrase frequency. This suggests that phrase frequency *is* represented in long-term memory, either as a single coherent representation or as individual episodes.

Recognition memory data provide a perspective on how speakers of a language map between linguistic material and a context. A canonical view of recognition is that, at test, speakers are given the linguistic content, the test items, and have to retrieve the experimental context in which they were experienced in order to endorse the items as old (Reder et al., 2000). The demands of a recognition task are therefore more comprehension-like than production-like. The other major memory task, recall, works in the opposite way. An act of recall starts with an existing temporal, discourse, or situational context representation ("recall all of the words *on the list you just saw*") and maps to the linguistic material that was experienced in this context (Criss, Aue, & Smith, 2011; Howard & Kahana, 2002). Recall is an explicit language generation task. In this respect, the demands of recall are more akin to production than comprehension. The current studies therefore examine phrase frequency effects in recall, rather than recognition, to provide a different perspective on the question of the source of such effects and what they tell us about phrasal representation.

Studies of language production demonstrate that frequent words and phrases are easier to say. Word and phrase frequency effects are apparent in a number of production measures including faster onset times (Janssen & Barber, 2012) and shorter word durations in frequent phrases (Arnon & Cohen Priva, 2013; Bannard & Matthews, 2008). Janssen and Barber assessed whether phrase frequency as measured by hits on the Google search engine predicted how easily speakers provided modified noun phrase picture descriptions like *blue car* or *red house* and noun-noun pairs like *bus car* in Spanish as well as noun-adjective pairs in French. They measured speech onset latencies as a function of phrase frequency, the frequency of the first word, and the frequency of the second word in each pair. When Janssen and Barber controlled for word frequency, phrase frequency explained the speedup in speech onset latencies, showing that high frequency phrases are easier to produce. Generally, the higher the phrase frequency, the earlier speakers began talking. Because they found phrase frequency effects, Janssen and Barber argued that phrases are stored holistically and that these representations lack a relationship between the component words and the phrase.

The results of Janssen and Barber were surprising because a previous study by Alario, Costa, and Caramazza (2002) had identified separable contributions of adjective and noun frequency to speech onset latencies, where high frequency adjectives and nouns sped up noun phrase production. Janssen and Barber argued that the results of Alario et al. could have also been due to variations in phrase frequency confounded with word frequency, as high frequency phrases tend to be made up of high frequency words, which have well-known frequency effects.

Additional evidence from child production data corroborates the hypothesis that the production system retrieves multiword units, perhaps in addition to individual words. Bannard and Matthews (2008) used a phrase imitation task in which children repeated phrases that an experimenter said to them. Children made fewer errors, and took less time to produce the overlapping words, when repeating more common phrases (e.g. "a drink of milk") than less common ones that shared the same first three words (e.g. "a drink of tea"). This suggests that long-term memory for multiword sequences has an effect on children's language production.

Theories of language production have not had a great deal to say about the production of phrases, with the possible exception of idiomatic phrases. The notion of a superlemma referred to earlier was developed by Sprenger et al. (2006) to allow for the model of Levelt, Roelofs, and Meyer (1999) to be able to produce idiomatic phrases. For non-idiomatic or compositional phrases, models have not assumed the existence of stored representations of multiword sequences (MacKay, 1982, is an exception in this respect). Because of the need for the production system to be able to assemble completely novel phrases (e.g. "an ugly beauty" cited by Chang, Dell, & Bock, 2006), models have emphasized that structural frames (e.g. adjective-noun) are retrieved, and then individual words UGLY and BEAUTY are retrieved and linked to slots in the frame (e.g. Chang et al., 2006; Dell, 1986; Dell, Oppenheim, & Kittredge, 2008; Garrett, 1975). Finding that production processes are sensitive to phrase frequency (e.g. Bannard & Matthews, 2008; Janssen & Barber, 2012) forces an amendment to these models.

To better understand phrase frequency effects, we consider the task of immediate free recall, which is an episodic memory task that engages the production system. We ask how phrase frequency supports retrieval for production. We will contrast phrase recall performance with recall of individual words. The first experiment (Experiment 1) explores the effects of word frequency on single-word (noun) recall, while Experiment 2 and Experiment 3 examine the influence of phrase frequency on recall of adjective-noun phrases.

Experiment 1

Frequency effects on free recall of nouns

The purpose of Experiment 1 is to examine whether a set of single words that show strong frequency effects in recognition in favor of the low frequency items (Balota, Burgess, Cortese, & Adams, 2002; Jacobs et al., 2016) exhibit similar frequency effects in a free recall task. Some studies have found no effect of frequency on recall (Clark & Burchett, 1994; MacLeod & Kampe, 1996; Hulme, Stuart, Brown, & Morin, 2003), while others have found an advantage for high frequency words (Balota & Neely, 1980; Criss et al., 2011).

When the words that we test for recall here were tested in yes-no recognition, the frequency effects were dramatic: the most common words had hit rates 35% lower than the least common words (Jacobs et al., 2016, Fig. 6). We expect, based on the prior literature, to find a very different effect for recall, and we will use this contrast to generate predictions for the effects of phrase frequency on phrase recall.

Methods

Materials

Study items were those used in Experiment 3 of Jacobs et al. (2016, Table A3). These items consisted of 88 nouns taken from Balota et al. (2002) that varied continuously in frequency in the Google 1T n-gram corpus (Brantz & Franz, 2006), a compilation of over 1 trillion words of text from the internet. The lowest frequency items include *parasol*, *sleuth*, and *crevice; car*, *book*, and *world* are examples at the high end and are reproduced in Appendix A.

Participants

Thirty individuals from the University of Illinois paid subject pool received $8 for participating. All were native English speakers who acquired no language other than English before the age of 5.

Procedure

Participants carried out an immediate free recall test of four 22-word lists. Each was made aware prior to list study that immediate written recall would take place. Each participant saw a unique ordering of 88 nouns that were randomly assigned to four lists, with the additional constraint that each list contained 11 high and 11 low frequency words. Study order was randomized within the list. Every word was presented at the center of the computer screen for 1 second, followed by a 1 second inter-stimulus interval before the presentation of the next item.

After the end of the presentation of each list, the computer presented a prompt for participants to start recalling the words they studied on a piece of paper with 22 spaces for each list. The prompt said, "Please fill in as many of the words as you can remember in any order you would like. Please try to recall as many words as you can." After acknowledging the instructions, the screen displayed a countdown showing the remaining amount of time to recall that list (5 min per list was allotted). At the end the five-minute recall period, participants could initiate study of the next list when they wished to by pressing a key.

Results

Every word that participants wrote down was entered as a data point for analysis. If participants wrote down an item that had appeared on an earlier list, that item was considered an intrusion and excluded from analysis. If items were misspelled but sufficiently similar to be identified as another item on the list (e.g. "alter" for "altar" or "yach", "yaght" and "yatch" for "yacht"), that item was included. Items that were not on any list that participants studied were not considered in the analysis.

To analyze the effect of word frequency on word recall, we constructed a logit mixed model of whether each item that participants studied was recalled or not as a function of (log transformed) word frequency and study order. Study order was entered as a quadratic variable to account for the bow-shaped serial position curve. This curve represents the memory effects of primacy and recency that are often seen in free recall (e.g. Anderson & Bower, 1972; Freebody & Anderson, 1986; May & Sande, 1982). Random effects of participant and item on the intercept and a random effect of participant on word frequency were included in the model.

Word frequency was not a significant predictor of the likelihood of the recall of a word. These results are summarized in Table 1 and plotted below in Fig. 1.

Discussion

The results of this study replicate prior findings of no high-frequency word advantage in the immediate recall of unrelated lists of nouns. As seen in Fig. 1, recall at the item level sits between 20% and 75% across the entire frequency range. Apparently, the strength of associations from episodic context to items does not reflect the commonness of the words. By itself, this null result does not have strong implications concerning the nature of lexical storage and retrieval. As we will see, however, the findings of Experiment 1, known effects of word frequency on recognition memory, and the effects of phrase frequency on recall that we will report in Experiments 2 and 3 will provide useful constraints on a model of the representation and retrieval of words and phrases.

Interim discussion

The results of Experiment 1 demonstrate that high frequency words are not necessarily better recalled than low frequency words. While this is in line with a number of previous studies that have not found an effect of word frequency on free recall, the pattern of results here differs from the expected pattern known to occur in less memory-focused language production tasks – when speakers are asked to name pictures, they are faster and more fluent in using high-frequency words (Dell, 1990; Jescheniak & Levelt, 1994; Kittredge, Dell, Verkuilen, & Schwartz, 2008).

The lack of a frequency effect on single-word recall can be explained if we consider the nature of recall. Recall can be conceptualized as a two-step process. First, recalling a word may involve mapping from the person's representation of the list of items they studied (hereafter known as the *episodic context*) to the word's semantic and/or syntactic representation, which is more formally known as the lemma. Second, once this representation is retrieved, the speaker must use the spoken or written production system to output the word.

The research mentioned above, taken together, specifically shows that it is the process of converting the lemma into speech or writing that is strongly sensitive to word frequency, rather than the retrieval of the lemma itself (Dell, 1990; Jescheniak & Levelt, 1994; Kittredge et al., 2008). The output of this conversion is ultimately phonological in nature, because speakers produce a sequence of sounds, letters, or characters. In unimpaired speakers this sensitivity is largely revealed in response time, rather than accuracy. Even though the word "wizard" is not particularly common relative to a word like "tree", when one has retrieved the lemma WIZARD, typical speakers accurately produce the word. By contrast, differences in production accuracy due to frequency only typically emerge in impaired populations (e.g. Kittredge et al., 2008), with one exception. Difficulty in phonological form retrieval during production is largely restricted to extremely low frequency words such "hemoglobin" or "ambergris." Difficulties retrieving these words often manifest as tip of the tongue states,

Table 1
Effect of word frequency on likelihood of noun recall.

	Estimate	SE	t
(Intercept)	−0.16	0.11	−1.40
(Log) word frequency	−0.05	0.05	−1.02
Study order (quadratic)	−0.0003	0.0002	−1.52

Note: Significance at $|t| > 2.00$.

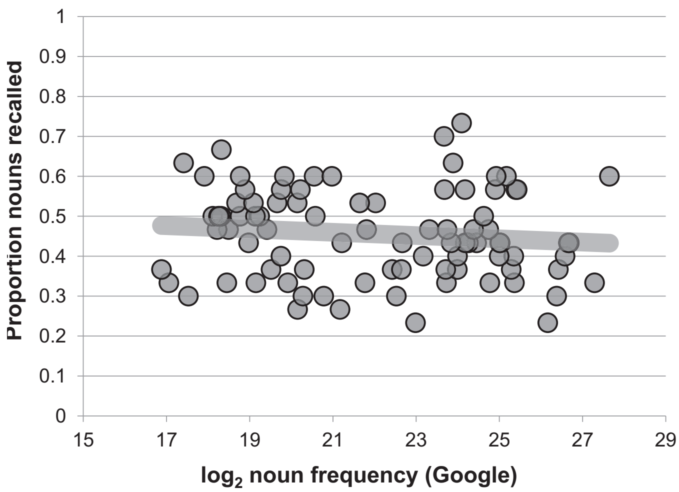

Fig. 1. Effect of word frequency on free noun recall, Experiment 1. More common nouns like *tree* are recalled just as often as less common nouns like *wizard*.

where the sounds corresponding to the word cannot be retrieved from what the speaker means to say (e.g. Brown & McNeill, 1966; Harley & Brown, 1998; Rubin, 1975).

Given these considerations, word frequency should not impact the production component of a typical untimed free recall task in which the words are known to the participant. Thus, if Experiment 1 had shown a substantial word frequency effect in free recall, it would have demonstrated frequency sensitivity in the link from episodic context to lexicon. Given that Experiment 1 and others (e.g. Dunlap & Dunlap, 1979; Ozubko & Joordens, 2007) have found no such word frequency effect on free recall, we tentatively conclude that word frequency is not a powerful influence on the episodic retrieval of a word.

Should phrase frequency then also not matter in free recall? One potential mechanism underlying the episodic retrieval of phrases is that phrase production benefits from pattern completion, otherwise known as *redintegration*. During redintegration, long-term memory associations between components of a to-be-recalled item help to fill in the gaps in memory when not all components are initially retrieved (Horowitz & Maneils, 1972; Schweickert, 1993). Phrase recall importantly differs from word recall in that phrases, unlike most words, are systematically composed of meaningful components (i.e. words). To the extent that free recall is driven by the retrieval of meaning (Hills, Jones, & Todd, 2012), one would expect systematic incomplete or partial recall in which some words are correct but not others. In such a case redintegration would mean that recall of some words of a phrase may help a speaker retrieve the other words. This process of redintegration may be sensitive to phrase frequency.

To see how phrase frequency might matter in recall, let us be more specific about redintegration in the recall of an adjective-noun phrase such as "alcoholic beverages." Assume that when the retrieval process starts, there is some probability p that at least one of the words is retrieved. Then, as recall continues, the remaining word may or may not be retrieved as well. We express the likelihood of successful retrieval of the other word, given that one of the words has been recalled, as the conditional probability, q. That is, q is the probability of recall of both words, given that at least one word is recalled. The three possible outcomes for a phrase (no recall, only one word is recalled, and both words are recalled) and their relation to p and q are illustrated in Fig. 2.

One can use this simple model to derive expectations about the role of phrase frequency in recall. Would high phrase-frequency aid initial recall, i.e. parameter p of the process? If we assume that initial recall is driven largely by the strength of the episodic associations from the list context to the language system and that these associations are not sensitive to frequency, as we claimed for single-word recall, then we do not expect a consistent effect of frequency on this parameter. Because phrases, however, are systematically composed of meaningful parts, retrieval from long-term memory representations may take advantage of connections between these components via a redintegration process. If so, we would expect more common phrases to be associated with *complete* recall, that is, to have a larger value of q (complete recall given some recall). We will postpone a consideration of specific mechanisms for such a process until we gather new data.

To test these proposals, Experiments 2 and 3 use recall tasks structured similarly to Experiment 1, except that the stimuli are meaningful adjective-noun phrases. In Experiment 2a, participants are presented with adjective-noun phrases designed to vary only in phrase frequency. After receiving the last phrase in a study session, participants must recall the phrases by writing them down. Experiment 2 b is a replication, except that participants are told to recall the individual words. This change was implemented in order to see whether the phrasal organization at output influences any phrase frequency effects. In Experiment 3, we sought to see whether the results of Experiment 2 generalized to another set of phrases and a different procedure in which study time per phrase was determined by the participants rather than being experimenter controlled. The results of these experiments allow us to test whether phrase frequency effects arise at initial recall of a phrase (complete or incomplete), or only after a participant has already recalled one of the words of a phrase.

Experiment 2a

Frequency effects on immediate free recall of adjective-noun phrases

If phrases are processed and remembered just as big words, then we expect phrase recall to be unaffected by phrase frequency, as seen in Experiment 1 with individual words. Hence, Experiment 2 looks at the effect of phrase frequency on free recall of phrases. Critically, phrase recall is prone to errors that single words cannot generate: parts of phrases can be recalled. We can capture this by estimating the two parameters that we outlined earlier, probability of some recall (p) and probability of complete recall given some recall (q).

Methods

Materials

Phrases from this experiment were a subset of the 112 phrases used in Experiment 3 of Jacobs et al. (2016). These phrases were taken from the spoken portion of the Corpus of Contemporary American English from 2009 to 2012 (COCA; Davies, 2008). This portion of the corpus contained approximately 17 million words. Stimuli include items such as "critical condition", "horrible mistake", and "impossible dream." To ensure that our assessment of the influence of phrase frequency on recall was not the result of any confounding between frequency and compositionality or concreteness, we conducted a norming study on Qualtrics in which University of Illinois undergraduates rated the items along several dimensions.

In this norming study, 30 participants were presented with phrases one at a time and, for each, provided responses to a number of questions on a five-point Likert scale from "Strongly Disagree" to "Strongly Agree". First, familiarity with the component words of each phrase and the phrase itself was assessed; participants answered whether they knew the meanings of, for example, the word "impossible", "dream", and the phrase "impossible

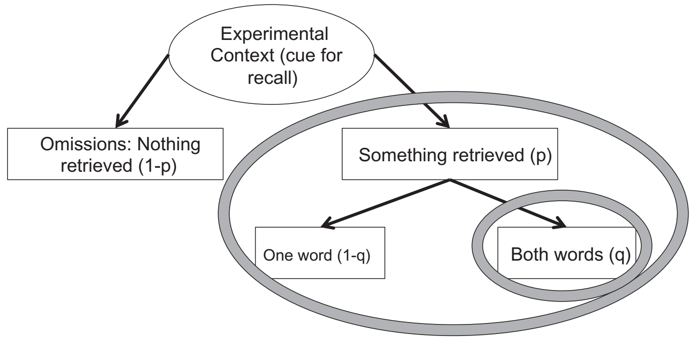

Fig. 2. In order to identify whether phrase frequency has an effect on the recall of phrases, two parameters can be estimated. First is the *p* parameter, which measures the likelihood of recalling at least one part of the phrase or the whole phrase (that is, either "alcoholic", or "beverages", or "alcoholic beverages") versus recalling nothing about a phrase. Second is the *q* parameter, which measures the conditional likelihood of recalling the entire phrase (i.e. "alcoholic beverages") given that something from the phrase (*p*) has been recalled.

dream." Then, to rate the imageability of the phrase, participants rated whether they could easily picture what this phrase describes. Finally, as a measure of compositionality, participants rated whether "impossible dream" had the same meaning as a dream that is impossible. Ratings were averaged across all participants and then centered and scaled with respect to all items for inclusion in the analyses. In the final stimulus set, phrases were restricted to just those where the average imageability and compositionality scores fell within a narrow range in order to decorrelate imageability and compositionality from phrase frequency ($r = 0.11$, $t(70) = 0.89$, $p =$ n.s. for imageability; $r = -0.14$, $t(70) = -1.19$, $p =$ n.s., for compositionality). Additionally, we verified that noun and adjective frequencies were not correlated with phrase frequency ($\rho = 0.11$, $t(70) = 0.95$, $p =$ n.s. and $\rho = 0.11$, $t(70) = 0.83582$, $p =$ n.s., respectively). After these requirements were met, 72 phrases remained. These stimuli are available in Appendix B.

Participants

In the norming study described above, 30 undergraduate students from the University of Illinois were recruited from the course credit subject pool. All participants were native speakers of English who acquired no other language before the age of 5. Each participant received one hour of credit for participation.

For the memory component of this study, a different set of 40 undergraduate students recruited from the University of Illinois course credit subject pool participated in this experiment with the same qualifications as the norming study. Each person received one hour of credit for participation in the experiment.

Procedure

The 72 items were broken into 4 lists and were randomly populated in the same way as in Experiment 1. Each list contained an equal number of high and low frequency phrases. For each list, participants studied 18 phrases for 1.5 s each followed by a 1 second inter-trial interval. After studying the 18th phrase, participants were told, "Try to write down as many of the phrases as you can remember. If you cannot remember both of the words from a phrase, but just one of the words, then write that down instead." Participants were given 5 min to complete recall of each list, again with a countdown informing them about how much time was left. If participants finished ahead of time, they waited until the timer finished before beginning study of the next list.

Results

Scoring

Each recalled item was categorized for whether the adjective was correctly recalled, whether the noun was correctly recalled, or both, as well as in what position in the recall list participants recalled the whole phrase or only part of the phrase. As before, items that could be identified as the target based on misspelling were included as correctly recalled in the analysis.

Phrase recall can be conceptualized as a two-stage process (e.g. Schweickert, 1993), which is summarized graphically in terms of the parameters p and q in Fig. 2. Participant responses on each individual trial were coded in terms of these whether participants had recalled at least one word ($p = 0$ or 1), and if they had recalled at least one word, whether they had recalled just one or both ($q = 0$ or 1). We then performed a sequential logistic regression analysis (e.g. Fox, 1997), fitting independent binary logistic models to each of the two stages. This tells us about the effect of phrase frequency on the likelihood of recalling anything from a phrase (first analysis) and the likelihood of partial versus complete recall (second analysis) respectively (that is, the p and q parameters).

Mixed effects logistic regression models were built to test for the effect of phrase frequency on recall, specifically on the likelihood of some recall (p) and the likelihood of complete-given-some recall (q). To account for as much of the variance as possible, we also included quadratic study order and concreteness on the likelihood of first some (p) and then complete-given-some (q) recall. Random effects were the participant-level random intercepts and random slopes of phrase frequency, with random intercepts by item.

Similar to the pattern of results in Experiment 1, where word frequency did not influence single word recall for nouns, we found that phrase frequency did not influence the likelihood at least one word of a phrase being recalled (the p parameter). Concrete phrases like "private plane", however, were more likely to be recalled at least in part than more abstract phrases like "critical condition." This occurred in spite of the relatively narrow range of concreteness values. Additionally, there was the expected effect of serial position, as seen in the significant quadratic study order term. These results are summarized below in Table 2.

The results with parameter q were different. There was a significant positive relationship between phrase frequency and the likelihood of the phrase being recalled in its entirety (given recall of at least one word) – high phrase frequency helped participants produce both words from studied phrases. More concrete phrases were also more likely to be recalled in their entirety. These results are summarized below in Table 3. Both results are plotted in Fig. 3 below. Note that the estimates of the *q* parameter for each phrase are typically higher than the estimates for *p* because *q* is conditioned on some recall, as we explained earlier. That is, although many phrases are not recalled at all, those that are, are comparatively often recalled in full.

Table 2
Effect of phrase frequency on parameter p, the recall of adjective-noun phrases, COCA stimuli, Experiment 2a. More common phrases are as likely to be recalled at least in part as less common phrases, but concrete phrases are more likely to be recalled.

	Estimate	SE	t	
(Intercept)	−0.41	0.11	−3.55	***
Study order (quadratic)	2.22	0.19	11.59	***
(Log) phrase frequency	0.01	0.10	0.12	
Phrase concreteness	0.32	0.10	3.17	**

Note: Significance at |t| > 2.00.

Table 3
Effect of phrase frequency on parameter q, the complete versus incomplete recall of adjective-noun phrases, COCA stimuli, Experiment 2a. More common phrases are more likely to be recalled in their entirety than less common phrases.

	Estimate	SE	T	
(Intercept)	1.92	0.21	9.28	***
Study order (quadratic)	1.21	0.39	3.10	**
(Log) phrase frequency	0.34	0.16	2.03	*
Phrase concreteness	0.50	0.15	3.38	***

Note: Significance at |t| > 2.00.

Experiment 2b

Experiment 2b was a replication of 2a with a change to recall instructions, emphasizing recall of words, rather than recall of phrases. Participants studied the same phrases as in Experiment 2a, but were told to write down as many of the individual words as they could remember.

Although the primary goal of Experiment 2b is to replicate Experiment 2a, Experiment 2b also tests whether phrase frequency matters even when phrases are not being overtly produced. For example, it might be the case that phrase frequency influences the process only when a noun-phrase structural frame (Dell, 1986; Garrett, 1975) is assembled for production, with the frequency of the phrase bound up with the slots in that frame. If frequency still matters without overt phrase production, we could perhaps speak of incidental activation of the co-occurrence of words within a phrase during recall. In any event, we will examine the role of phrase frequency in the same manner that we did for Experiment 2a.

Methods

Materials

Materials were the same as those from Experiment 2a.

Participants

40 participants from the University of Illinois course credit or paid subject pool took part in this experiment. All participants were native English speakers who acquired no other language before the age of 5.

Procedure

The study procedure of this experiment was identical to that of Experiment 2a. The recall phase differed in the instructions given to the participants about the nature of their responses after study. Participants were told, “You are going to see a series of two-word phrases presented on the screen. While they are two words presented together, we want you to remember each of the individual words separately because you will be asked to write down the individual words on separate lines from memory. If you remember both words from a phrase, write each word on a separate line.” After participants began the test phase, they again had 5 min to recall as many of the words as possible by writing their answers on sheets of paper with provided spaces. At the end of the five-minute recall period, participants pressed a key to begin the next study-test phase.

Results

Experiment 2b replicated the effects of Experiment 2a. Participants wrote down at least one word from a phrase as often across all frequency ranges (results for parameter p in Fig. 4 and Table 4), but were significantly more likely to recall both words from high frequency phrases given recall of at least one word (results for q in Fig. 4 and Table 5). Words from more concrete phrases were more likely to be recalled (p) and were more likely to be recalled if their phrasal mate had been recalled (q). Finally, as before words that had occurred in phrases early or late in the list were better recalled than words from phrases in the middle of the list.

Discussion

Experiment 2b combined features of Experiment 1 and Experiment 2a by assessing whether the recall of individual words from

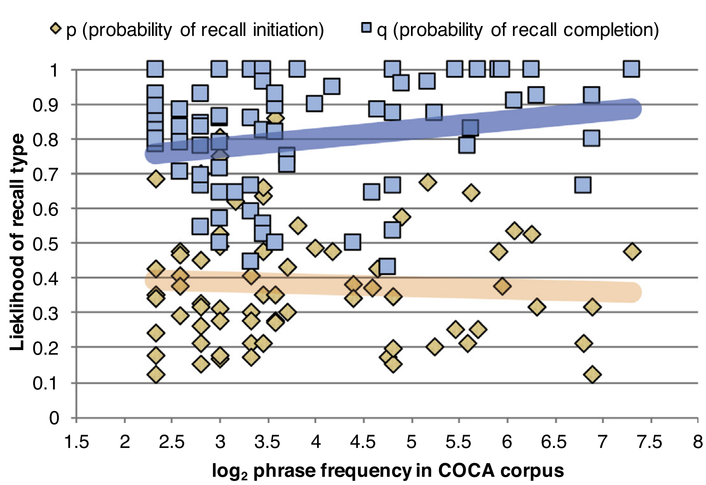

Fig. 3. Effect of phrase frequency on the recall of adjective-noun phrases from COCA, Experiment 2a. More common phrases are more likely to be recalled in their entirety (blue squares) than less common phrases, but all are equally likely to be recalled at least in part (orange diamonds). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

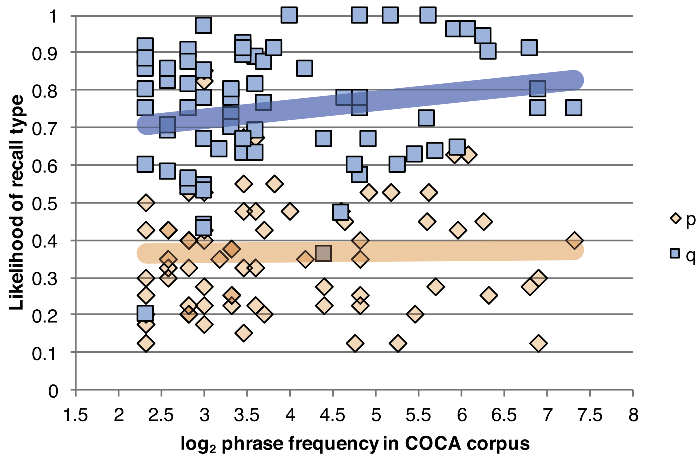

Fig. 4. Effect of phrase frequency on the recall of individual words from adjective-noun phrases from COCA, Experiment 2b. More common phrases are more likely to lead to both words being recalled (blue squares) than less common phrases, but all are equally likely to be recalled to some extent (orange diamonds). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
Effect of phrase frequency on parameter p, the probability of recall of any of the words from adjective-noun phrases, COCA stimuli, Experiment 2b.

	Estimate	SE	t	
(Intercept)	−3.86	0.62	−6.14	***
Study order (quadratic)	0.01	0.001	8.60	***
(Log) phrase frequency	0.03	0.03	1.04	
Phrase concreteness	0.31	0.07	4.78	***

Note: Significance at |t| > 2.00.

Table 5
Effect of phrase frequency on parameter q, the complete versus incomplete recall of adjective-noun phrases, COCA stimuli, Experiment 2b. More common phrases are more likely to be recalled in their entirety than less common phrases.

	Estimate	SE	t	
(Intercept)	−0.99	0.10	−9.86	***
Study order (quadratic)	0.40	0.04	9.10	***
(Log) phrase frequency	0.13	0.04	2.76	**
Phrase concreteness	0.23	0.05	5.13	***

Note: Significance at |t| > 2.00.

concrete, compositional phrases was affected by phrase frequency. Like words, compositional phrases were recalled at least in part (parameter p) equally well at all levels of the frequency range, which is similar to the effect found in Experiment 1 for individual words. Experiment 2a found that once something had been retrieved from a phrase, though, the phrase was more likely to be completed if it was a high frequency phrase than if it was a low frequency phrase (parameter q). Experiment 2b replicated these results, demonstrating that long-term memory representations of high frequency phrases are useful for pattern completion, in that the retrieval of one word in a phrase facilitates the retrieval of the other word in a phrase.

There was one difference in the results of Experiments 2a and 2b (see Fig. 5). When participants recalled an item, they recalled both words of the phrase as opposed to just one word in Experiment 2a on average 79.8% of the time, while in Experiment 2b this value was only 74.4%. In a paired *t*-test comparing p and q parameters of the two experiments, although the effect of frequency on q was similar in the experiments, the q values themselves are significantly lower in Experiment 2b ($t(71) = -2.27$, $p = .013$). At the same time, participants were just as likely to recall an item in whole or in part (parameter p) in Experiment 2b as 2a ($t(71) = -0.92$, $p = 0.185$).

It is striking that even when the task is not to recall phrases, but instead individual words, the influence of phrase frequency on word recall is similar to its effect in phrase recall. This suggests that phrasal organization in long-term memory is the driving force behind phrase frequency effects in free recall. Furthermore, there is a dissociation between p and q in how influential the instructions are. Telling participants to write down single words as opposed to phrases affects the likelihood of participants writing down both words of a phrase when they recall an item (q), but does not influence the likelihood of them recalling at least one word from that item (p).

In summary, even though participants were asked to recall individual words, the task demands did not prevent them from recalling both words from a phrase. This is consistent with the finding that in single word recall, participants often attempt to recall temporally contiguous or semantically related words at the same time (Gruenewald & Lockhead, 1980; Lohnas & Kahana, 2014; Sederberg, Howard, & Kahana, 2008; Unsworth, Brewer, & Spillers, 2014; Wixted & Rohrer, 1994). The results of Experiment 2 are consistent with the idea that the initial recall of a word or phrase is insensitive to phrase frequency, but that once a part of the phrase has been recalled, phrase frequency becomes an important catalyst in recalling an entire phrase.

Experiment 3

Frequency effects on self-paced study and free recall of adjective-noun phrases

Experiment 2 demonstrated that phrase frequency can affect aspects of phrase recall, especially during the process of completing recall of an entire phrase. Experiment 3 aimed to replicate and extend the phrase frequency effects of Experiment 2 in a recall paradigm where participants can pace their own study and where the materials differ from prior materials by having a wider range of concreteness scores. While it was less clear what would happen in more natural materials with the likelihood of the initial recall of any given phrase (the p parameter), the analysis of the q parameter representing the likelihood of redintegration remains the critical analysis. If phrase frequency influences the likelihood of the complete recall of a phrase, then Experiment 3 should replicate the

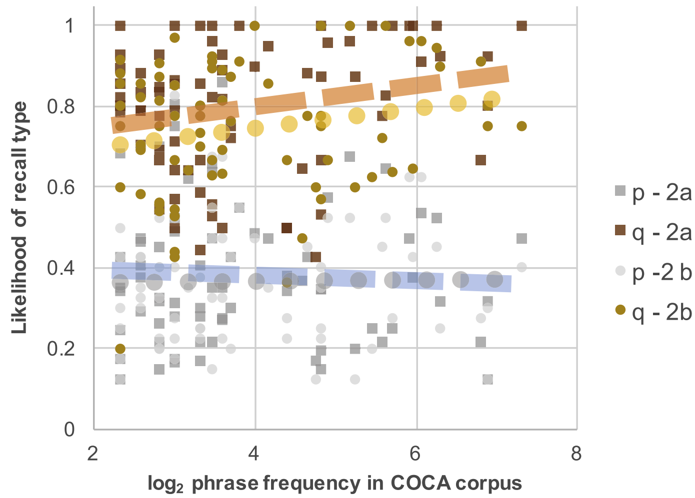

Fig. 5. Comparison of Experiments 2a and 2b. The effect of phrase frequency is similar in both Experiment 2a and 2b for both the likelihood of remembering either one or two words (p) and for the likelihood of remembering two words when at least one word was recalled (q). Participants are less likely to recall phrases completely in Experiment 2b, the experiment in which they were prompted to only write down words rather than phrases.

effects of Experiment 2 on the q parameter, with high frequency phrases being more likely to be recalled in their entirety than low frequency phrases.

Methods

Materials

Experiment 3 used the 52 phrases from Experiment 1 of Jacobs et al. (2016) as stimuli such as "alcoholic beverages" and "psychic nephew". These items varied in their phrase frequency, which was decorrelated by design from adjective frequency, noun frequency, and both word lengths, but which somewhat confounded concreteness with phrase frequency. These stimuli are reproduced in Appendix C. Phrase frequency and concreteness were correlated ($\rho = .49$), which we account for in later analyses by performing likelihood ratio tests.

Participants

Seventy-nine undergraduate participants were recruited from the University of Illinois course credit subject pool. All participants were native speakers of English who acquired no language other than English before the age of 5. Each person received one hour of credit for participation in the experiment.

Procedure

Each participant did two study-test blocks of 26 phrases each that were randomly populated in the same way as in Experiments 1 and 2. Phrases were presented at the center of the screen until participants pressed the space bar to continue on to the next phrase followed by a one second inter-item interval. After pressing a key to complete study on the 26th item, the test phase began. Participants were told, "Try to write down as many of the phrases as you can remember. If you cannot remember both of the words from a phrase, you can write down just one of the words." Participants were given 10 min per list to recall as many of the items as they could remember by writing their answers on sheets of paper with provided spaces. Participants waited the entire interval before beginning the second study-test phase.

Results

Random effects were structured in the same way as Experiment 2. Fixed effects of interest included how long a participant studied each item in log seconds, quadratic study order, the concreteness of each phrase taken from the norms of Jacobs et al. (2016), and the log frequency of the phrase. Because concreteness and study time were somewhat confounded with the variable of interest (phrase frequency), we performed likelihood ratio tests for whether including phrase frequency in the model explained variance over above that explained by a model containing only study time, concreteness, and study order. When the likelihood ratio test revealed that adding frequency gave a significant improvement in fit, we included phrase frequency in the final model.

We were also interested in whether participants studied phrases more when they were infrequent, which could weaken or eliminate any phrase frequency effects on memory (though see the *laboring in vain effect*; Nelson & Leonesio, 1988). Participants indeed studied less common phrases for longer periods of time ($B = -0.013$, $t = -4.57$, $p < .001$), in line with similar frequency-related processing fluency gains in studies of language comprehension (Arnon & Snider, 2010; Siyanova-Chanturia et al., 2011; Smith & Levy, 2013). In spite of this, the extra labor on the uncommon phrases did not eliminate the tendency for common phrases to be better recalled, as we show below (see Fig. 6).

The first analysis focuses on the likelihood of recalling at least one word from a phrase (parameter p). The analysis showed that length of time the participants studied an item, the order of the item in a list, and its concreteness all influenced the likelihood of a phrase being recalled. Furthermore, the inclusion of phrase frequency improved model fit beyond these control variables ($\chi^2(1) = 8.46$, $p < .01$). The model assessing the effect of phrase frequency on the p parameter is summarized below in Table 6.

Focusing on the q parameter, phrase frequency importantly continued to have an effect on recall performance. The model containing phrase frequency, concreteness, study time and study order was a better fit than a model that contained all of these factors other than phrase frequency in a likelihood ratio test ($\chi^2(1) = 6.99$, $p < .01$). Even when controlling for these other factors, higher frequency phrases like "alcoholic beverages" were more likely than lower frequency phrases like "psychic nephew" to be recalled as wholes. Phrases studied for longer as well as those with higher concreteness ratings were more also associated with higher values of q. These results are summarized below in Table 7.

Below are plotted the likelihood of some recall (p) and complete recall given any recall (q) as a function of phrase frequency in Fig. 7.

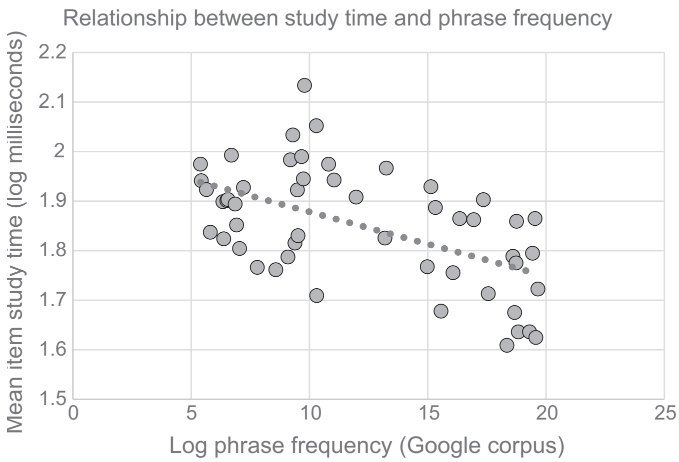

Fig. 6. Effect of phrase frequency on study time. More common phrases are studied for less time.

Table 6
Effect of variable on the p parameter of the recall of adjective-noun phrases, Google stimuli. Concrete phrases, those that are studied longer, and higher frequency phrases are associated with a higher value of p.

	Estimate	SE	t	
(Intercept)	−0.38	0.13	−2.82	**
(Log) phrase frequency	0.35	0.12	3.03	**
(Log) study time	0.51	0.07	7.67	***
Study order (quadratic)	0.40	0.05	8.51	***
Phrase concreteness	0.30	0.12	2.48	*

Note: Significance at |t| > 2.00.

Table 7
Effect of phrase frequency on the complete versus incomplete recall of adjective-noun phrases, Google stimuli. More common phrases are more likely to be recalled in their entirety than less common phrases.

	Estimate	SE	t	
(Intercept)	2.14	0.22	9.70	***
(Log) phrase frequency	0.57	0.21	2.66	**
(Log) study time	0.29	0.12	2.37	*
Study order (quadratic)	0.07	0.10	0.70	
Phrase concreteness	0.44	0.21	2.04	*

Discussion

Experiment 3 demonstrated that phrase frequency has a strong influence on the likelihood of a phrase being recalled in its entirety, given some recall, replicating the findings of Experiment 2. Experiment 3′s replication of Experiments 2a and 2b's positive phrase frequency effect on q solidifies a conclusion that redintegrative processes drive the reproduction of phrases from memory. A possible explanation of this effect is that phrasal representations consist of their constituent words, with some kind of link, such as a direct association or a chunk node joining them. In any event, the phrases are not atomic. In the general discussion we consider these results in concert with other findings regarding phrase and word frequency effects in recall and recognition.

General discussion

Frequent linguistic units facilitate fluent language production. High frequency words are produced more quickly (Ellis, 2002; Forster & Chambers, 1973; Gahl, 2008) and are less prone to errors (Dell, 1990; Nozari, Kittredge, Dell, & Schwartz, 2010). Production is a component of verbal free recall, so we can ask whether common linguistic units benefit in recall as well. Although the present study did not consistently find that phrase frequency contributed to the probability that at least one word of a studied adjective-noun phrase is recalled, the facilitative effect of frequency did show up as a greater likelihood of complete phrase recall (as opposed to partial recall). We characterized this finding as phrase frequency consistently affecting one parameter (q), but not the other (p), of a two-stage description of phrase recall.

The results of our experiments fit nicely with other investigations of the recall of adjective-noun phrases (e.g. Bower, 1969; Horowitz & Maneils, 1972; Paivio, Khan, & Begg, 2000). In a seminal study, Horowitz and Maneils (1972) tested for the influence of idiomaticity on the free recall of such phrases. Phrases were either idiomatic expressions like *sour grapes*, meaningful (compositional) adjective-noun phrases like *green grapes*, or anomalous like *deep grapes*. Participants were told to write down as many phrases from memory as possible as part of a free recall task. As in the present study, Horowitz and Manelis were interested in whether the different kinds of phrases were more likely to be recalled as wholes, as opposed to partially. They found that although idiomatic phrases had a very strong tendency to be recalled as wholes, the compositional and even the anomalous phrases tended to be recalled as wholes too. This effect demonstrates the influence of redintegrative processes during phrase retrieval.

Redintegration refers to a process of pattern completion using information from long-term memory (Horowitz & Maneils, 1972; Schweickert, 1993; Thorn, Gathercole, & Frankish, 2005; Hulme et al., 1997). We propose that specifically in phrase recall, the representations of words that are retrieved during language production cue one another to the extent that they have often co-occurred. We believe that the representation at which this cuing takes place is not at the level of the word form (i.e. the actual sounds of the word) but instead at either the abstract syntactic representation of the word, which in production theory is called the *lemma*, or the higher "lexical concept" level which is a semantic, but word-specific representation (Levelt et al., 1999). Our data do not allow us to choose between lemma and lexical-concept levels as the locus of the redintegration. Thus, as we develop our model below, when we refer to the "words" of a phrase and "word nodes", please recognize that we are referring to higher-level (non-phonological) representations without a further commitment to their precise level or nature.

The following proposal outlines a model of our findings concerning the role of frequency in word and phrase recall. This model also explains the effects of word and phrase frequency in recognition memory, particularly the data from Jacobs et al. (2016), who investigated phrase and word frequency effects using similar materials to those employed here.

The main challenge for a model of memory for linguistic material such as words and phrases is the fact that frequency effects appear to behave quite differently in recall and recognition. In particular, such a model must first be able to explain the well-known finding that more common words have considerably worse discriminability in recognition (Glanzer & Adams, 1985; Jacobs et al., 2016, Experiment 3a), but, in single-word free recall, word frequency often has little impact on performance (our Experiment 1; Dunlap & Dunlap, 1979; Ozubko & Joordens, 2007). The results for phrases are even more complex, with frequency mattering for some aspects of each memory task, but not for other aspects. High frequency phrases are more likely to be recalled in their entirety once recall of a single word has been initiated (the consistent effects of phrase frequency on the q parameter), but there is a relative lack of phrase frequency effects on the p parameter, (Experiments 2a, 2b). In recognition, high-frequency phrases garner more

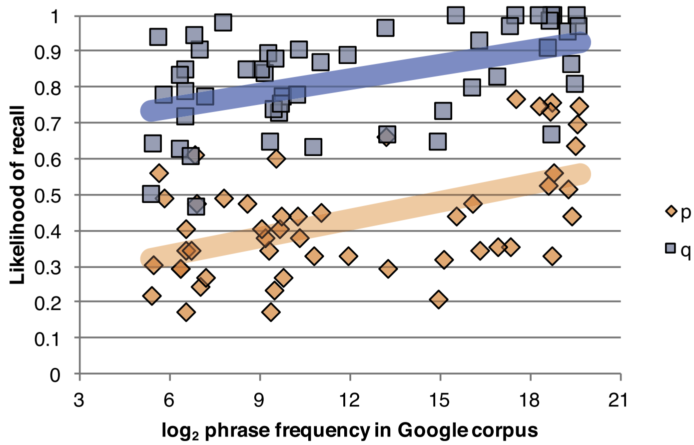

Fig. 7. Effect of phrase frequency on parameters p and q in the recall of adjective-noun phrases, Google stimuli, Experiment 3. More common phrases are more likely to be recalled in whole or in part than uncommon phrases, and are more likely to be recalled in their entirety given that at least one of the words was recalled (blue squares) than less common phrases. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 8
Pattern of results that the model must be able to account for and proposed mechanisms.

Experimental result	Mechanism
Recall	
Low frequency words and high frequency words are equally likely to be recalled	Links from episodic context to lexical or semantic representations of words are independent of frequency
Low frequency phrases and high frequency phrases are under some conditions, equally likely to be recalled at least in part	Links from episodic context to lexical or semantic representations of words (and therefore phrases) are independent of frequency
High frequency phrases are more likely than low frequency phrases to be completed once one word has been recalled	Associations between the words within the lexical-semantic system are stronger in high frequency phrases
Concrete phrases are easier to recall than abstract phrases	Concrete phrases have more active features, so the associations between a new episode and a concrete phrase is stronger
Recognition (Jacobs et al., 2016)	
Low frequency words are better discriminated than high frequency words	Studied high frequency words suffer from more interference from prior episodes
High frequency phrases get more "yes" responses regardless of whether they were studied or not (a bias)	Associations between the words within the lexical-semantic system are stronger in high frequency phrases, contributing to greater familiarity
High and low frequency phrases are equally well discriminated	There are many more episodes sharing a word in a phrase than the whole phrase. Thus, interference from other phrase episodes is minimal
Low frequency words facilitate phrase discrimination	Compositional phrases access word episodes, so high frequency words within phrases generate more interference just as they do in recognition for single words
Concrete phrases are better discriminated than abstract phrases	Concrete phrases have more active features, so the associations between a new episode and a concrete phrase are stronger

"yes" responses during recognition tasks (Jacobs et al., Experiments 1 and 2), but phrase-frequency does not impact actual discriminability. Instead, the frequency that impacts phrase discriminability in recognition is *word* frequency, specifically the frequency of the noun in adjective noun phrases (Jacobs et al., 2016).

Finally, it is worth noting a property of phrase memory that appears to work similarly in recall and recognition: Concrete phrases are better remembered (Experiment 2a, 2b, and 3 for recall and Experiment 1 in Jacobs et al.; Kusyszyn & Paivio, 1966; Paivio et al., 2000). In Table 8, we summarize the pattern of results from the present word and phrase recall studies as well as the word and phrase recognition studies of Jacobs et al. (2016), and provide a brief characterization of how each effect is explained in the model that we detail below.

The model we propose combines features of language production models with prominent models of episodic memory (e.g. Howard & Kahana, 2002; Reder et al., 2000). An episodic memory is a link between features of the context and features of an item. The context represents the participant's surroundings, her internal state, and her conception of the task. During the study of a list, the set of context features will gradually change, but we assume that a great many will remain constant and thus represent the "list". The study item has features that represent properties of particular studied words and phrases. These features arise from processing the linguistic material using the lexical-semantic system that is used for language production and comprehension. Item features would include semantic and syntactic properties of the item, as well as possible sensory-motor features that are called to mind by processing the meaning of the material. This means that linguistic and conceptual properties of "psychic", "nephew", as well as emergent conceptual properties about "psychic nephew" (e.g. "I have too many eccentric relatives!") are all potential features. In our model, we will represent the collection of features associated with a word, such as *cat*, by a single node. But, this is a shorthand for the word's many lexical-semantic (as opposed to phonological) properties.

Recall and recognition are handled differently by the model but make use of the same architecture. The start point for recall is always the context, and the goal of recall is to use the context to retrieve linguistic material associated with it; that is, speakers are attempting to produce a word or a phrase. Recognition, instead, starts with the linguistic material as a cue. The recognition process succeeds (or generates a hit) when the linguistic input cues retrieval of the crucial experimental episode in which the material was studied. At the same time, the recognition process is influenced by the familiarity of the linguistic information, so unstudied material that is very familiar can trigger a "yes" response.

We assume that studied words and phrases are features of stored episodes. An episode is a node connecting a representation of the episodic context and the lexical/ semantic representations of

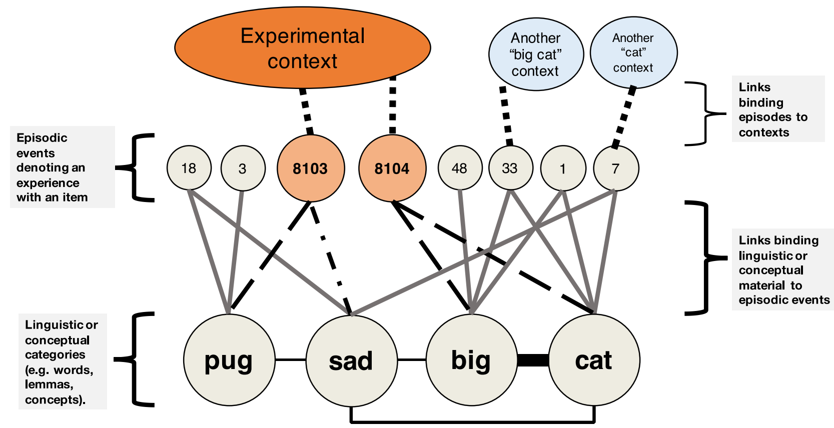

Fig. 8. Diagram of the model of phrase frequency effects in recall and recognition. Frequent phrases are associated with more episodes and with stronger associations within the lexical-semantic network.

the linguistic material. The strength of the link between the context and the linguistic material is not assumed to reflect frequency of usage of the linguistic material. But material that is more concrete is assumed to contain more features and thus to have a potentially richer linkage.

More frequent words and phrases are assumed to be linked to more episodes. In addition, more common words have stronger connections to their phonological forms (e.g. Dell, 1990; Jescheniak & Levelt, 1994; Kittredge et al., 2008; Nozari et al., 2010). Phrases that have been heard or produced before include a link (or node, e.g. MacKay, 1982) connecting the lexical/semantic representations of their component words, with more common phrases having stronger connections.

These assumptions are illustrated in Fig. 8. The pool of episodes contains numbered nodes that represent experiences. Each episode is therefore an instance, or exemplar, of a particular (potentially linguistic) category or combination of categories. In the figure, for example, Node 33 denotes a memory involving something big and something about cats, such as the phrase *big cat*. Likewise, Node 18 indicates an encounter involving something sad and a pug, potentially a *sad pug*. The lines linking lexical-semantic information to episodic events do not reflect frequency, but potentially concreteness and the activation or amount of attention devoted to the words.

The episodes are not all attached to exactly the same context features, since experiments unfold over time. A participant's experience of the beginning of the experiment may be different from the end of it, for example. So, episodes should be able to be bound to different parts of a context. To illustrate this, the *big* and *cat* episode (Node 8104) connects to a different part of the context than the *sad* and *pug* episode (Node 8103), as different information may have been salient at time points 8103 and 8104.

Below we outline how these assumptions explain the word and phrase frequency effects in recall and recognition.

Recall

Low frequency words and high frequency words are equally likely to be recalled

This suggests that the long-term memory encoding process, that is, the linkage between each episode and the words that participated in it, is largely independent of any frequency information that is stored with the representations of words in the lexical/semantic network. Lexical frequency is presented in the lexical semantic system, but it is most strongly felt in the mapping from semantic/syntactic representations to phonological forms. During a recall test of familiar words by unimpaired speakers, the sensitive component of the mapping does not generate any appreciable error.

High frequency phrases are more likely than low frequency phrases to be completed once one word has been recalled

Because participants are capable of recalling phrases incompletely, we assume that episodes include links separately to each word in the phrase. Recall that Experiments 2a, 2b, and 3 demonstrated that phrase frequency effects arose at the level of the completion of a phrase given that recall had been initiated (that is, the q parameter value increased as phrase frequency increased). In light of these results, Fig. 8 links individual words to episodes. When two words are experienced at the same time, these words attach to the same episode. This architecture allows for participants to not necessarily recall both words from a phrase. Note that there are more episodes linking *big* and *cat* together (Nodes 1, 33, and 8104) than episodes linking *sad* and *pug* (Nodes 18 and 8103). Participants must use the context to guide what items they recall: this top-down search requires also locating episodes that are associated with the experiment only and not unrelated episodes. Starting with a given context effectively eliminates all other instances of a phrase (that is, all other *big cat* episodes) during recall. Phrase frequency effects like we saw at the level of phrase completion require phrase frequency to be encoded elsewhere.

We can relate the process of retrieving both words from a phrase as being similar to spreading activation. When speakers retrieve one word, they are able to retrieve a related word more easily because words associated with previous material in long-term memory become active. In the phrase case, the next word in a phrase becomes easier to retrieve. In the architecture of this model, we represent the capacity for spreading activation between two words as solid bars connecting the words within the word layer in Fig. 8. The more often two words occur together, the stronger the connection between them, and the more likely that both words will be retrieved once one has been produced.

Phrase frequency does not always affect initial recall of words

Recall that Experiment 1 found no effect of word frequency on word recall success and we explained this by assuming that the strength of the episodic links to the words is largely independent of lexical frequency. For a non-idiomatic phrase, we assume that its episodic representation consists of links from its words to the episode. That is, there is no phrase node (e.g. a "psychic nephew" concept) that is linked to the episode. Instead, the individual words are jointly linked to an episode. Given this, we expect little effect of phrase frequency on the first stage of recall, when words are initially retrieved from the context. This is what we found in Experiments 2a and 2b, in which phrase frequency did not influence the p parameter. We note that there was an effect of phrase frequency on p in Experiment 3, though. It is possible that the longer study times used in Experiment 3 allowed, during memory encoding, for the assumed stronger associations between the words of common phrases to increase the activations of both words of the phrase to an extent greater than would occur for less common phrases. If it is further assumed that the level of activation of each word during encoding contributes to the resulting strength of the association between the word and the episodic context, then an effect of phrase frequency on p might be expected.

Concrete phrases are easier to recall than abstract phrases

Concrete and imageable words and phrases are typically much easier to understand, produce, recognize, recall, and learn. In every experiment in this study, concreteness influenced the likelihood of the initial retrieval of a phrase (the p parameter) as well as the likelihood of the completion of a phrase given initial retrieval (the q parameter). We propose that the number of features associated with a studied word or phrase determines the strength of the link between a new episode and the item. Concrete words and phrases (e.g. *alcoholic beverages*) have a number of perceptual features that more abstract words and phrases (e.g. *psychic nephew*) do not, such as texture and color (Grondin, Lupker, & McRae, 2009; Marslen-Wilson & Warren, 1994; Plaut & Shallice, 1993; Vinson & Vigliocco, 2008; Wiemer-Hastings & Xu, 2005). These richer sensory representations make the initial retrieval of a word or phrase easier than for more abstract words and phrases. For example, the advantage for concrete words in sentence production leads concrete words to be mentioned first in sentences (e.g. Bock & Warren, 1985).

Recognition

Any satisfactory model of phrase memory must be able to account for frequency effects in recognition memory in addition to recall. Low frequency words like *pug* are much more easily discriminated in recognition than high frequency words like *cat*. Phrase recognition differs: Jacobs et al. (2016) found that participants discriminated high and low frequency *phrases* equally well, even though there was a strong bias to say that they had studied high frequency phrases like *alcoholic beverages* but not low frequency ones like *psychic nephew*. They did find that words within phrases impacted discriminability, such that participants best remembered phrases that contained low frequency nouns like *handsome wizard*. In light of these results, the model must not allow for low frequency phrases to be better discriminated than high frequency phrases, but phrases with rare words should be better recognized.

How does recognition memory take place in this model? We can conceptualize recognition as the inverse of recall. Instead of going from the context to retrieving linguistic content, participants start from linguistic content in order to retrieve a context, which participants verify as part of the experiment or not. When participants read the words on a computer screen, they retrieve the episodes associated with those words (some of which overlap because of previous co-occurrence). Then, participants search within those episodes to determine whether that episode was part of the experiment.

Low frequency words are better discriminated than high frequency words

Studied low frequency words like *pug* are easier to recognize because they have fewer episodes than common words like *cat*, so participants find the experimental episode with less competition from other episodes. Unstudied low frequency words are easier to recognize because it is also easier to verify that no studied episode exists. In this respect, the model captures well-known effects captured by a number of other models (e.g. Hintzman, 1988; Mandler, 1980; Reder et al., 2000).

High frequency phrases get more "yes" responses regardless of whether they were studied or not (a bias)

The bias originates from the same spreading activation-like mechanism that facilitates the completion of more common phrases in free recall. Once one word has been processed, associated words that co-occur regularly activate each other. So, once a participant has read a word like *big*, the word *cat* receives greater activation than before and is therefore easier to process. This more fluent processing leads to the illusion of the phrase having been studied – regardless of whether it was studied or not, and leads to a bias among participants to say that they have studied high frequency phrases.

High and low frequency phrases are equally well discriminated

Generally speaking, phrases are much less frequent than the words that compose them. If we assume compositional phrase representations, then recognition requires searching through episodes bound to individual words, potentially in addition to episodes bound to phrases. Following from the account in Jacobs et al. (2016), we propose that the relative contribution of phrase frequency to episodic search will be much less influential than word frequency due to the existence of fewer phrase episodes, so discriminability of phrases will not be sensitive to their frequency.

Low frequency words facilitate phrase discrimination

Since the number of episodes associated with at least one word within a phrase is much larger than the number of episodes containing the whole phrase, test phrases containing high frequency words will have many episodic memories of those high frequency words that can impede search through episodic memory. This leads to an advantage for recognition of phrases containing uncommon words (for similar proposals, see Jacobs et al., 2016; Reder et al., 2000; Malmberg, Steyvers, Stephens, & Shiffrin, 2002).

Concrete phrases are better discriminated than abstract phrases

By the same mechanism as we proposed in free recall, more concrete phrases have stronger links to an episode because they have more features. When a concrete phrase is presented during recognition, the link between that phrase and the critical episode is stronger, which leads to greater discriminability of concrete phrases.

Conclusion

We have examined word and phrase frequency effects in free recall. As is also true for such effects in recognition, the results are not straightforward. Words and phrases are not necessarily better recalled when they are more frequent. But in the case of phrases, there is a clear benefit for high frequency phrases for complete, as opposed to partial recall. We presented an informal model of these data and corresponding data in word and phrase recognition that put effects of word and phrase frequency in two locations in the cognitive system – within the lexical-semantic system that is responsible for language production and comprehension, and in the system that creates episodic memories based on the features that the lexical semantic system generates.

Acknowledgments

The first author was supported by an NSF Graduate Research Fellowship and a Beckman Graduate Research Fellowship. This project was funded in part by NIH-R01 DC008774 and NIH-R01 HD086099.

A. Experiment 1 materials

High frequency nouns	$\log_2$ frequency	High frequency nouns	$\log_2$ frequency	Low frequency nouns	$\log_2$ frequency	Low frequency nouns	$\log_2$ frequency
nation	24.74	truck	23.84	anvil	18.11	tripod	20.31
library	25.33	mouth	24.18	vulture	17.52	beggar	18.31
home	28.37	cow	22.02	pecan	18.21	jaguar	18.88
valley	22.64	radio	25.28	owl	20.54	wharf	18.77
chicken	23.16	plane	23.72	sleuth	17.05	flea	20.21
sun	24.38	wheel	23.71	parasol	16.87	flask	19.14
garden	24.25	bottle	23.31	valet	19.13	keg	18.24
rose	23.67	tree	25.01	altar	20.77	harp	19.91
palace	21.54	street	25.01	isle	19.24	vine	20.15
floor	24.99	engine	25.34	otter	18.49	urn	19.83
town	25.42	picture	26.16	dungeon	19.65	crevice	17.41
baby	25.17	bread	22.98	gourd	17.90	dwarf	20.35
field	26.41	pool	24.77	sequin	18.26	vase	20.57
road	25.36	key	26.37	eel	18.76	galaxy	21.21
cloud	22.42	cup	23.99	lily	20.14	boar	18.68
father	24.92	book	27.28	cavern	19.75	yacht	20.97
hotel	26.58	jacket	22.66	gem	21.17	tunic	18.52
snake	21.56	beach	24.62	cobra	19.41	tablet	21.77
village	23.99	market	26.66	loft	20.27	olive	21.80
world	27.64	cat	24.09	plum	19.75	banjo	19.51
dress	23.68	king	23.74	bonnet	18.97	silo	18.32
car	26.68	bear	23.88	wizard	21.64	monsoon	19.09
college	25.39	stream	24.90	apa	22.52		
kitchen	24.17	ball	24.45	lass	18.45		

B. Experiment 2 materials

Adjective	Noun	$\log_2$ adjective frequency	$\log_2$ noun frequency	$\log_2$ phrase frequency	Imageability	Compositionality
effective	treatment	10.22	10.61	2.32	4.31	4.38
impossible	dream	10.01	10.55	2.32	3.90	3.93
open	relationship	10.92	11.80	2.32	4.48	3.38
poor	credit	10.74	11.21	2.32	4.00	3.38
sad	truth	10.36	11.67	2.32	4.39	3.83
serious	nature	11.94	10.33	2.32	3.68	2.52
similar	incident	10.61	10.19	2.32	4.00	4.07
fair	deal	10.46	11.95	2.58	4.07	4.14
funny	feeling	11.21	10.23	2.58	4.21	2.54
heavy	heart	10.18	11.92	2.58	4.21	2.31
major	bank	11.89	10.95	2.58	3.89	3.15
physical	violence	10.31	11.34	2.58	4.55	4.38
british	actor	10.55	10.65	2.81	4.24	4.38

(continued on next page)

Experiment 2 materials (*continued*)

Adjective	Noun	log_2 adjective frequency	log_2 noun frequency	log_2 phrase frequency	Imageability	Compositionality
necessary	step	10.15	10.39	2.81	4.14	3.83
normal	behavior	10.76	10.54	2.81	4.24	4.07
positive	test	10.75	10.65	2.81	3.90	3.14
safe	space	10.96	11.08	2.81	4.07	4.14
successful	mission	10.71	10.92	2.81	4.00	4.28
violent	weather	10.01	11.16	2.81	4.52	4.00
actual	cost	10.16	10.77	3	4.28	3.31
available	flight	10.64	10.59	3	4.31	4.28
easy	solution	11.05	10.26	3	4.31	4.31
fresh	blood	10.68	11.29	3	4.21	3.28
iraqi	freedom	10.97	10.6	3	3.57	2.68
quick	action	10.9	11.29	3	3.97	3.97
senior	officer	11.06	10.81	3	4.14	3.17
white	neighborhood	11.69	10.34	3	4.38	2.97
international	agreement	11.62	10.35	3.17	3.86	4.00
full	pcture	11.59	11.64	3.32	4.55	2.43
likely	suspect	10.8	10.05	3.32	4.38	3.41
lucky	break	10.17	11.64	3.32	3.93	2.86
strong	opinion	11.85	10.98	3.32	4.28	3.79
terrible	accident	10.87	10.6	3.32	4.59	3.97
clear	winner	11.55	10.08	3.46	4.00	2.79
current	governor	10.82	11.81	3.46	4.34	4.24
fine	art	11.35	10.01	3.46	4.04	2.54
military	background	11.84	10.04	3.46	4.17	3.14
super	model	10.41	10.29	3.46	4.69	2.28
emotional	response	10.35	11	3.58	4.31	4.17
horrible	mistake	10.13	10.67	3.58	4.48	4.28
sexual	act	10.56	10.9	3.58	4.28	4.55
short	film	10.98	11.50	3.58	4.45	4.31
commercial	success	10.38	10.79	3.7	4.07	2.76
global	recession	10.89	10.27	3.7	4.00	4.14
healthy	weight	10.35	10.61	3.81	4.41	3.83
guilty	pleasure	11.19	10.36	4	4.24	3.10
innocent	victim	10.04	10.63	4.17	4.24	4.28
extraordinary	amount	10.19	11.3	4.39	3.97	4.00
personal	choice	11.55	11.16	4.39	4.14	4.03
independent	investigation	10.15	11.74	4.58	3.83	3.72
beautiful	song	11.86	11.33	4.64	4.29	4.41
significant	progress	10.87	10.45	4.75	4.21	4.00
amazing	experience	11.57	11.61	4.81	4.28	4.48
crrect	answer	10.22	11.47	4.81	4.59	4.59
enormous	pressure	10.15	11.21	4.81	4.21	3.48
powerful	message	10.73	11.84	4.81	4.21	4.17
private	plane	11.51	11.19	4.91	4.69	4.31
single	parent	11.57	10.07	5.17	4.69	3.86
Close	Attention	10.91	11.71	5.25	4.28	2.36
main	course	10.58	10.88	5.46	4.59	2.59
recent	study	11.09	10.72	5.58	4.28	4.14
tough	love	11.89	11.72	5.61	4.00	3.52
early	age	11.72	11.62	5.7	4.59	3.45
low	income	10.1	10.61	5.91	4.48	4.00
social	network	11.61	10.88	5.95	4.41	2.93
supreme	leader	10.93	11.5	6.07	4.28	3.76
hot	seat	11.14	10.69	6.25	3.86	1.86
critical	condition	10.67	10.03	6.3	4.34	4.10
wrong	direction	10.98	10.65	6.8	4.41	4.28
popular	vote	10.82	11.43	6.88	4.24	2.97
regular	basis	10.29	10.06	6.89	4.21	3.10
common	ground	10.80	11.64	7.3	4.07	2.29

C. Experiment 3 materials

Phrases from Jacobs et al. (2016) used in Experiment 3

	Phrase		log_2 phrase frequency	log_2 adjective frequency	log_2 noun frequency
	Adjective	Noun			
Low frequency phrases	simultaneous	transduction	5.39	21.48	19.70
	downstream	subcontractors	5.42	21.57	19.98
	naughty	tot	5.64	21.88	20.14
	abandoned	arena	5.80	22.36	22.71
	accompanying	visions	6.33	22.31	20.91
	packaged	hunts	6.37	21.72	19.43
	chrome	throttle	6.50	21.53	20.22
	optimum	staining	6.50	21.69	20.45
	flaming	bounds	6.55	19.67	21.65
	predominant	organ	6.70	20.15	22.39
	psychic	nephew	6.85	21.10	20.43
	transgenic	allele	6.91	20.17	19.88
	inhaled	compounds	7.04	19.60	22.64
	programmable	fuse	7.20	20.82	20.55
	sleek	fleece	7.79	20.91	20.68
	piercing	headache	8.57	21.04	21.47
	metropolitan	zones	9.09	21.69	22.61
	decadent	era	9.19	19.22	23.28
	commanding	brigade	9.29	20.23	19.95
	distinct	affinity	9.38	23.20	21.07
	routine	expressions	9.48	23.32	22.56
	untreated	asthma	9.51	20.18	22.02
	painful	consciousness	9.66	22.27	22.50
	tangled	headset	9.74	19.34	21.38
	intense	cultivation	9.79	22.76	20.93
	oerennial	grasslands	10.29	20.41	19.03
High frequency phrases	thick	bundles	10.30	23.50	20.49
	vibrant	acidity	10.80	21.61	19.28
	polynomial	curves	11.04	21.11	22.09
	cherished	traditions	11.97	19.88	22.31
	passionate	embrace	13.18	21.58	21.71
	accumulated	surplus	13.24	21.61	22.18
	conditional	expectation	14.97	21.83	21.80
	relentless	pursuit	15.13	19.84	21.84
	unsecured	tenant	15.32	21.64	21.81
	roman	numerals	15.56	20.28	19.25
	interior	decoration	16.06	23.48	21.41
	contaminated	soils	16.35	21.81	21.83
	undue	hardship	16.94	20.31	20.60
	outer	shell	17.35	22.83	23.43
	dining	hall	17.55	23.44	23.09
	mashed	potatoes	18.34	19.37	21.71
	respiratory	tract	18.59	22.01	21.93
	cystic	fibrosis	18.67	19.37	19.85
	cerebral	palsy	18.73	20.98	19.39
	monoclonal	antibody	18.75	19.99	22.03
	bald	eagle	18.82	22.00	21.54
	nitric	oxide	19.30	19.75	21.74
	myocardial	infarction	19.42	20.37	19.93
	coronary	artery	19.53	21.29	21.35
	acoholic	beverages	19.56	21.34	21.55
	rheumatoid	arthritis	19.65	19.93	21.79

References

Alario, F. X., Costa, A., & Caramazza, A. (2002). Frequency effects in noun phrase production: Implications for models of lexical access. *Language and Cognitive Processes, 17*, 299–319.

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review, 79*, 97–123.

Arnon, I., & Cohen Priva, U. (2013). More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech, 56*, 349–371.

Arnon, I., & Cohen Priva, U. (2014). Time and again: The changing effect of word and multiword frequency on phonetic duration for highly frequent sequences. *The Mental Lexicon, 9*, 377–400.
Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language, 62*, 67–82.
Baayen, R. H., Milin, P., Đurdević, D. F., Hendrix, P., & Marelli, M. (2010). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning.
Baayen, R. H., Hendrix, P., & Ramscar, M. (2013). Sidestepping the combinatorial explosion: An explanation of n-gram frequency effects based on naive discriminative learning. *Language and Speech, 56*, 329–347.
Balota, D. A., Burgess, G. C., Cortese, M. J., & Adams, D. R. (2002). The word-frequency mirror effect in young, old, and early-stage Alzheimer's disease: Evidence for two processes in episodic recognition performance. *Journal of Memory and Language, 46*, 199–226.
Balota, D. A., & Neely, J. H. (1980). Test-expectancy and word-frequency effects in recall and recognition. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 576–587.
Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science, 19*, 241–248.
Bock, J. K., & Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition, 21*, 47–67.
Bower, G. H. (1969). Chunks as interference units in free recall. *Journal of Verbal Learning and Verbal Behavior, 8*, 610–613.
Brantz, T., & Franz, A. (2006). The google web 1T 5-gram corpus. *Linguistic Data Consortium*. https://research.googleblog.com/2006/08/all-our-n-gram-are-belong-to-you.html, https://catalog.ldc.upenn.edu/LDC2006T13.
Brown, R., & McNeill, D. (1966). The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior, 5*, 325–337.
Bybee, J. L. (2006). From usage to grammar: The mind's response to repetition. *Language, 82*, 711–733.
Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review, 113*, 234–272.
Clark, S. E., & Burchett, R. E. (1994). Word frequency and list composition effects in associative recognition and recall. *Memory & Cognition, 22*, 55–62.
Copestake, A., Lambeau, F., Villavicencio, A., Bond, F., Baldwin, T., Sag, I., & Flickinger, D. (2002). Multiword expressions: Linguistic precision and reusability. In *Proceedings of the 3rd international conference on language resources and evaluation (LREC 2002)* (pp. 1–7).
Criss, A. H., Aue, W. R., & Smith, L. (2011). The effects of word frequency and context variability in cued recall. *Journal of Memory and Language, 64*, 119–132.
Davies, M. (2008). *The corpus of contemporary American English*. BYE, Brigham Young University.
Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 9*, 283–321.
Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes, 5*, 313–349.
Dell, G. S., Oppenheim, G. M., & Kittredge, A. K. (2008). Saying the right word at the right time: Syntagmatic and paradigmatic interference in sentence production. *Language and Cognitive Processes, 23*, 583–608.
Dunlap, G. L., & Dunlap, L. L. (1979). Manipulating the word frequency effect in free recall. *Memory & Cognition, 7*, 420–425.
Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition, 24*, 143–188.
Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior, 12*, 627–635.
Freebody, P., & Anderson, R. C. (1986). Serial position and rated importance in the recall of text. *Discourse Processes, 9*, 31–36.
Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language, 84*, 474–496.
Garrett, M. F. (1975). The analysis of sentence production. *Psychology of Learning and Motivation, 9*, 133–177.
Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition, 13*, 8–20.
Grondin, R., Lupker, S. J., & McRae, K. (2009). Shared features dominate semantic richness effects for concrete concepts. *Journal of Memory and Language, 60*, 1–19.
Gruenewald, P. J., & Lockhead, G. R. (1980). The free recall of category examples. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 225–240.
Harley, T. A., & Brown, H. E. (1998). What causes a tip-of-the-tongue state? Evidence for lexical neighbourhood effects in speech production. *British Journal of Psychology, 89*, 151–174.
Hills, T. T., Jones, M. N., & Todd, P. M. (2012). Optimal foraging in semantic memory. *Psychological Review, 119*, 431–440.
Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95*, 528–551.
Horowitz, L. M., & Maneils, L. (1972). Toward a theory of redintegrative memory: Adjective-noun phrases. *Psychology of Learning and Motivation, 6*, 193–224.
Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology, 46*, 269–299.
Hulme, C., Stuart, G., Brown, G. D., & Morin, C. (2003). High-and low-frequency words are recalled equally well in alternating lists: Evidence for associative effects in serial recall. *Journal of Memory and Language, 49*, 500–518.
Jacobs, C. L., Dell, G. S., Benjamin, A. S., & Bannard, C. (2016). Part and whole linguistic experience affect recognition memory for multiword sequences. *Journal of Memory and Language, 87*, 38–58.
Janssen, N., & Barber, H. A. (2012). Phrase frequency effects in language production. *PLoS ONE, 7*, e33202.
Jescheniak, J. D., & Levelt, W. J. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 824–843.
Kittredge, A. K., Dell, G. S., Verkuilen, J., & Schwartz, M. F. (2008). Where is the effect of frequency in word production? Insights from aphasic picture-naming errors. *Cognitive Neuropsychology, 25*, 463–492.
Kusyszyn, I., & Paivio, A. (1966). Transition probability, word order, and noun abstractness in the learning of adjective-noun paired associates. *Journal of Experimental Psychology, 71*, 800–805.
Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). Multiple perspectives on word production. *Behavioral and Brain Sciences, 22*, 61–69.
Lohnas, L. J., & Kahana, M. J. (2014). Compound cuing in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 12–24.
MacKay, D. G. (1982). The problems of flexibility, fluency, and speed–accuracy trade-off in skilled behavior. *Psychological Review, 89*, 483–506.
MacLeod, C. M., & Kampe, K. E. (1996). Word frequency effects on recall, recognition, and word fragment completion tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 132–142.
Malmberg, K. J., Steyvers, M., Stephens, J. D., & Shiffrin, R. M. (2002). Feature frequency effects in recognition memory. *Memory & Cognition, 30*, 607–613.
Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87*, 252–271.
Marslen-Wilson, W., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review, 101*, 653–674.
May, R. B., & Sande, G. N. (1982). Encoding expectancies and word frequency in recall and recognition. *The American Journal of Psychology, 95*, 485–495.
Morgan, E., & Levy, R. (2016). Abstract knowledge versus direct experience in processing of binomial expressions. *Cognition, 157*, 384–402.
Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect". *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 676–686.
Nozari, N., Kittredge, A. K., Dell, G. S., & Schwartz, M. F. (2010). Naming and repetition in aphasia: Steps, routes, and frequency effects. *Journal of Memory and Language, 63*, 541–559.
Ozubko, J. D., & Joordens, S. (2007). The mixed truth about frequency effects on free recall: Effects of study list composition. *Psychonomic Bulletin & Review, 14*, 871–876.
Paivio, A., Khan, M., & Begg, I. (2000). Concreteness of relational effects on recall of adjective–noun pairs. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 54*, 149–160.
Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology, 10*, 377–500.
Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember–know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 294–320.
Rubin, D. C. (1975). Within word structure in the tip-of-the-tongue phenomenon. *Journal of Verbal Learning and Verbal Behavior, 14*, 392–397.
Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition, 21*, 168–175.
Sederberg, P. B., Howard, M. W., & Kahana, M. J. (2008). A context-based theory of recency and contiguity in free recall. *Psychological Review, 115*, 893–912.
Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*, 523–568.
Siyanova-Chanturia, A., Conklin, K., & Van Heuven, W. J. (2011). Seeing a phrase "time and again" matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 776–784.
Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition, 128*, 302–319.
Sprenger, S. A., Levelt, W. J., & Kempen, G. (2006). Lexical access during the production of idiomatic phrases. *Journal of Memory and Language, 54*, 161–184.
Thorn, A. S. C., Gathercole, S. E., & Frankish, C. R. (2005). Redintegration and the benefits of long-term knowledge in verbal short-term memory: An evaluation of Schweickert's (1993) multinomial processing tree model. *Cognitive Psychology, 50*, 133–158.
Unsworth, N., Brewer, G. A., & Spillers, G. J. (2014). Strategic search from long-term memory: An examination of semantic and autobiographical recall. *Memory, 22*, 687–699.
Vinson, D. P., & Vigliocco, G. (2008). Semantic feature production norms for a large set of objects and events. *Behavior Research Methods, 40*, 183–190.
Wiemer-Hastings, K., & Xu, X. (2005). Content differences for abstract and concrete concepts. *Cognitive Science, 29*, 719–736.
Wixted, J. T., & Rohrer, D. (1994). Analyzing the dynamics of free recall: An integrative review of the empirical literature. *Psychonomic Bulletin & Review, 1*, 89–106.