Measuring Mental Entrenchment of Phrases with Perceptual Identification, Familiarity Ratings, and Corpus Frequency Statistics

Catherine Caldwell-Harris (charris@bu.edu)
Department of Psychology, Boston University
64 Cummington St., Boston, MA 02215, USA

Jonathan Berant (jonatha6@post.tau.ac.il)
School of Computer Science, Tel-Aviv University
Tel-Aviv, 69978, Israel

Shimon Edelman (se37@cornell.edu)
Department of Psychology, Cornell University
Ithaca, NY 14853, USA

--

1 Introduction

Word recognition is the Petri dish of the cognitive sciences. The processes hypothesized to govern naming, identifying and evaluating words have shaped this field since its origin in the 1970s. Techniques to measure lexical processing are not just the back-bone of the typical experimental psychology laboratory, but are now routinely used by cognitive neuroscientists to study brain processing and increasingly by social and clinical psychologists (Eder, Hommel, and De Houwer 2007). Models developed to explain lexical processing have also aspired to be statements about the nature of human cognition (e.g., connectionist models, Plaut, McClelland, Seidenberg, and Patterson 1996). Words were convenient objects to study for cognitive psychologists because they are well-defined and their nature as alphabetic strings was a good fit to analysis with the computer programming languages of the 1970s and 1980s which excelled at string manipulation.

But are words actually the privileged unit of mental representation and processing that all of this scientific attention makes them out to be? Like a growing number of other language researchers, our answer is no (see, e.g., (Bybee and Hopper 2001; Wray 2002). We propose that the mental representations for lexical structures form a continuum, from word combinations which have fossilized into single units (nightclub) to those that both exist as independent units and yet have bonds, varying in tightness, with the words with which they frequently co-occur (Harris 1998).

The first line of support for this view is the simple observation that fluent speakers easily recognize the familiarity and cohesive quality of word combinations in their language. Examples in English include common noun compounds (*last year, brand new*), verb phrases (*cut down, get a hold of, faced with*) and other multi-word expressions such as common sayings and references to cultural concepts (*saved by the bell, speed of light*; Jackendoff 1995). These frequently co-occurring word sequences, referred to as collocations or multi-word sequences, have only recently been studied by psycholinguists, but several recentpublications have argued that language processing and human cognition need to be expanded (if not altered) to accommodate speakers' wide-ranging knowledge of common word combinations (e.g., Arnon and Snider 2010; Bannard and Matthews 2008).

Speakers' demonstrable facility with multiword expressions may be uncomfortable for many linguists and psycholinguists because of these fields' historical allegiance to parsimony in representation, as exemplified by the words and rules approach to language (Pinker 1999). This approach proposes that the building blocks of language are a set of basic units (words) and rules for combining them into larger structures.

Like the older proposal of "large words" as the way to explain idioms (Swinney and Cutler 1979), the words-and-rules approach specifies that non-compositional expressions are stored as unanalyzed wholes in the lexicon, while all compositional expressions, even highly frequent ones, are produced by combining words.

It seems that a shift in how these phenomena are discussed has occurred over the last decade. Historically, idioms were regarded as the main problematic entity for dictionary-style models of the mental lexicon. Some theorists noted that treating idioms as unanalyzed wholes or "large words" conflicted with the compositional pieces of idioms and their partial grammatical productivity (Gibbs and Nayak 1989). Recently, a greater concern is the question of the mental representation and processing of frequent multi-word-expressions. Idioms and non-compositionality are thus viewed as just one aspect of the larger phenomenon of multi-word-utterances.

1.1 Proposals for how to represent multi-word expressions

Intuitively, we can note that when an expression represents the kind of content that speakers want to say repeatedly and often, it becomes entrenched and gains some kind of unitized storage. Continuing to speak informally, unitization (of some type) and repeated use means that additional semantic connotations may attach themselves to the utterance. With historical changes in grammar and lexicon that can be successfully resisted by this high frequency unit, non-compositionality emerges.

Two broad categories of models are available to theorists who want to address the representational questions posed by multiword expressions that vary in frequency, as discussed by Arnon and Snider 2010; see also Snider and Arnon, this volume).

Frequency threshold approach: Phrases of sufficient frequency have independent representation as a way of making processing more efficient. Open questions are what counts as sufficient frequency or whether other factors play a role in establishing a multiword structure as a linked or unitized structure.

Continuous approach: Every instance of usage affects processing and representation. The continuous approach is an implication of adopting an emergentist or dynamical systems framework (Ellis and Larson-Freeman 2006, 2009; Elman 1995; MacWhinney 1999). It also assumes the usage-based approach to language developed by linguists working in cognitive grammar (Langacker 1987) and functionalist grammar (Beckner & Bybee 2009; Bybee and Hopper 2001). According to the usage-based hypothesis, each use of an expression influences its entrenchment and future processing (Tomasello 2003). The difference between more and less frequent is thus one of degree, rather than specifying whether the sequence is stored vs. computed.

1.2 Studying language processing as if frequency mattered

It has become common to assert that frequency effects on behavior are pervasive and appear in every aspect of language processing (Arppe, Gilquin, Glynn, Hilpert, and Zeschel 2010). Still, considerable knowledge gaps exist, along with theoretical disagreements about the extent of frequency effects and how to account for them. Frequency effects are studied by corpus linguists, psycholinguists, and computational modelers. Frequency variation across utterances is of course very salient to corpus linguists. A pressing concern for these researchers has been to link this variation to other types of data, such as determining whether the most frequent use of a polysemous word or grammatical sequence is also the prototype (Gilquin and Gries 2009). Arppe et al. (2010) have argued for the need to use convergent methods, combining corpus analysis with various types of elicited data, including behavioral experiments, noting that there is "little or no understanding of how results from these different types of data inform one another" (p. 7). A second challenge identified by Arppe et al. (2010) is to establish conventions for interpreting corpus findings as cues to psychological entrenchment. This point is echoed by Gilquin and Gries' (2009) review of studies that use both corpora and experiments. While researchers typically hope for convergence, divergent outcome using different methods doesn't necessarily invalidate either finding, because specific measures are sensitive to specific linguistic activities (reading, speaking, comparing, judging etc., see Divjak 2008).

Psycholinguists are interested in both theoretical issues such as testing the "words and rules" and usage-based models, as well as applied topics concerning frequency. A recent example is how native vs. non-native speakers vary in their histories of usage patterns. Non-native speakers may have too little experience to have built up language routines and multiword expressions and may thus rely on translating from their first language, with consequent errors and lack of native-like output.

Computational modelers face the conceptual challenge of how to implement frequency effects while meeting other modeling goals such as inducing grammatical and lexical regularities.

Our own investigation into frequency effects in multiword utterances is broadly inspired by emergentist approaches, and a specific computational model, ADIOS (Automatic DIstillation Of Structure; Solan, Horn, Ruppin, and Edelman 2005). A goal in designing AIDOS was to develop an unsupervised learning algorithm that could induce grammar from raw text. Collocations are inherently important for grammar in ADIOS, as the algorithm learns to represent not just the classical phrase structure constituents of grammars, but also the full range of multiword expressions, including those that show partial or complete productivity.

The method that ADIOS uses to induce grammatical structures from raw text is to search for recurring sequences of words. It can work on linguistic input of any size and quality, including short phrases such as found in children's speech or adults' speech to children. Given an input sequence, the algorithm scans its current corpus of utterances, seeking stored phrases that share a contiguous sequence of lexical items. For instance, the test sequence *I saw the news today* may pull out from the corpus the utterances *the news is good* and *I read the news online these days*, all aligned on the shared subsequence *the news*. If this shared subsequence represents a statistically significant collocation, it becomes unitized as a collocation.

To learn a grammar that can generate novel utterances rather than merely extracting regularities, the criterion that governs matching in ADIOS is relaxed to allow a local mismatch in the shared subsequence. By allowing local mismatches, the phrases *I heard the big news* and *I heard the latest news* would match. Allowing two phrases to match even if they diverge in part of the sequence means that units (lexical items or collocations) occupying the corresponding slots in the aligned phrases will be interchangeable in the context of the parent collocation. The result is the formation of equivalence classes, which are added to the growing grammar. This procedure is repeated recursively until no new collocations are found. The resulting grammar represents both the sequential order of lexical items and larger collocations.

ADIOS has proven to be effective at grammar induction (see Waterfall, Sandbank, Onnis, and Edelman 2010), but still unknown is how closely algorithms like AIDOS match behavioral data on frequency effects. It may may be premature at present to model behavioral data because of lack of knowledge about boundary conditions. That is, we don't know when frequency effects stop appearing, but ADIOS suggests some testable possibilities. An interesting aspect of ADIOS for frequency effects in multiword utterances is that a sequence can have low global usage frequency, yet stand out in a circumscribed set of contexts and become entrenched and unitized (Waterfall, Sandbank, Onnis, and Edelman 2010). In the current study we thus looked for frequency effects using low frequency collocations. ADIOS also predicts that fully compositional sequences are stored, such as *the latest news* or even *the news*. We thus also tested whether adjective+noun patterns like *her list* would show frequency effects. ADIOS assumes that efficiency of processing a sequence is a function of quantity of exposure for individual learners. This encouraged us to test whether speakers with different types of exposure to phrases would be differentially efficient at processing those phrases.

Below we describe two studies that present new evidence for the sensitivity of language users to the variations in frequency in multi-word utterances. Both studies use the same psycholinguistic technique, perceptual identification, a powerful method for measuring strength of representations that has thus far been underused to study processing of multiword utterances.

1.3 New types of frequency effects

As noted, it has become common to find frequency effects in a diversity of types of language processing (Arppe et al. 2010). Still unclear is whether these findings require embracing the continuous model, as predicted by usage-based hypothesis and emergentist approaches, or whether a frequency threshold approach is sufficient. Arnon and Snider (this volume) have provided evidence against the threshold approach by reporting frequency effects for more than two frequency groupings. For their response latency data, a continuous measure of frequency always had a superior fit than a dichotomous (low/high) measure, a finding consistent with continuous models. We want to push this a step farther by studying more of the frequency spectrum. Continuous models like ADIOS predict frequency effects for low frequency collocations, and even for sequences which are merely legal sequences, but wouldn't be called collocations, like *her list, size three*, and *some cans*. The second novel area of inquiry is to find frequency effects as a function of speakers' expertise. Speakers who have more experience with some specific expressions should perform more efficiently in identifying those expressions than speakers with less exposure/use of those word sequences.

1.4 The perceptual identification task

In a standard perceptual identification task, a stimulus is briefly displayed on a computer screen, typically for durations of 30 ms to 100, and then masked with a visual noise pattern, which disrupts continued processing (Carr, 1986; Ratcliff and McKoon 1997). The subjective experience of respondents may be of seeing an unknown word, but with a few practice trials participants feel they can guess and are often correct or close. Exposure durations are usually long enough so that stimuli can be consciously perceived, but short enough that response accuracy is below ceiling. Brief display and masking makes the recognition task difficult. The difficulty is reduced, and recognition enhanced, if participants can easily match the brief, degraded input to a representation in long term memory. The stronger the long term memory representation, the more accurate is identification. Classic phenomena using the perceptual identification task include the word superiority effect, in which observers can more easily identify a word than a nonword (Rumelhart and McClelland 1981). A variant of the standard task is to display multiple words sequentially (as done in Caldwell-Harris and Morris 2009), but the dependent measure remains the same, which is to identify some or all items in the perceptual display.

2 Frequency effects from collocations, to legal expressions to random word pairs

The growing interest in relating behavioral measures to results of corpus analysis (e.g., Gilquin and Gries 2009), and the public availability of high quality corpora such as the Corpus of Contemporary American English (COCA; Davies 2010), opens the door for pycholinguists to draw on a rich source of data about frequency effects in multiword utterances: data from existing experiments which can be reanalyzed by connecting performance data to corpora.

Caldwell-Harris and Morris (2009) identified a temporal illusion produced when observers perform perceptual identification on familiar word combinations. When the word combinations were highly frequent, but presented sequentially in reverse order (i.e., *code* followed by *zip*), observers report perceiving the familiar word pair *zip code*. For exposure durations ranging from 30 to 105 ms. for each word, observers spontaneously reversed word pairs such as *fees legal* and *step next*.

Report of words in familiar order persisted even when observers were informed that some words would be presented in reversed order and that it was important to report the order in which words appeared, even if this was not the familiar order. Participants reported that their subjective impression was that a reversed pair such as *fees legal* had been displayed in its canonical order (*legal fees*). This impression held up even for experienced observers such as laboratory assistants who were familiar with the words on the stimulus list. We will refer to these as reversal errors, but of course they are errors only from the standpoint that observers are not sensitive to order of word presentation, but are instead reporting the words in their most frequent order. As discussed further in that paper, this performance could be seen as optimal from a Bayesian perspective, since the prior probability of *card credit* as an independent two-word display is much lower than the probability of *credit card*.

The probability of making a reversal error was highest for high frequency collocations (*keep track, fan club*), next highest for low frequency collocation (*machine gun, any clues*), and next for adjective+noun combinations (*huge church, real skin*). Perceiving veridical order was highest for the random word pairs (*look fever, puppy hill*). The ability to correctly recognize the component words regardless of their order was strongly influenced by the frequency category of the word combination, and only minimally influenced by the frequency of the individual words in the string.

The data set included word pairs across the frequency range, from the highest word pairs in COCA (*thank you*) to those with low and zero frequency, meaning that this perceptual identification data deservers a deeper analysis. We pursue a more in-depth analysis here and link recognition accuracy to three types of familiarity ratings, and two frequency corpora, Google and COCA, thereby responding to corpus linguists' plea for more work linking up different types of frequency measures with different types of behavioral measures (Arppe et al. 2010; Gilquin and Gries 2009).

2.1 Description of the stimuli

The inclusion of random word pairs and the merely legal word pairs in the Caldwell-Harris and Morris (2009) data allow us to investigate whether frequency effects exist even for word pairs of absent and low frequency. Because the study was designed and administered from 1997-1999, collocation frequency was determined by a corpus consisting of electronic newsgroup postings used for the HAL project (Hyper Analog to Language, Lund and Burgess 1996; in 1997 Kevin Lund generously gave the authors a list of all word pairs that occurred more than 5 times in the HAL corpus). High-frequency word pairs had a mean frequency of 3700 in the 300 million word corpus, while low frequency pairs had a mean frequency of 20.6. Adjective + noun combinations were selected to be legal combinations but 0 frequency, and thus they did not exist in the HAL corpus, but were constructed to avoid violating semantic constraints, following Pustejovsky's (1995) description of semantic domains. The random word pairs were mainly noun-noun pairs which violated semantic domains, and could not easily be assimilated to an adjective + noun combination, or to any easily identifiable legal syntactic grouping, although considerable variation resulted. Stimuli, corpora frequencies, ratings and recognition accuracy appear in the Appendix.

2.2 Intercorrelations between familiarly and frequency

Google frequencies were obtained by placing quotation marks around each word pair. Frequencies ranged from a low of 74 for the random pair *weep job* to a high of 819,000,000 for the high frequency collocation *health care* (*thank you* was a close second in frequency; the high frequency of *health care* is probably an artifact of the heavy use of this phrase when Google frequencies were collected in July 2007). For COCA, a corpus of 410 million, frequencies ranged from 0 to 77,530 (*thank you* being the most frequent and *health care* the 4th most frequent of these word pairs, at 28,620). Log frequency was used in graphs and calculations.

Subjective familiarity ratings for the 160 word pairs were obtained from 22 undergraduates. Raters used a 5-point scale extending from very unfamiliar to very familiar. Raters were additionally given the option of evaluating a phrase as "does not make sense." Phrases so rated were scored as 0, resulting in a familiarity scale ranging from 0 to 5.

The correlation between Google and COCA frequencies was high (r=.93). Correlations obtained separately just on the collocations and legal pairs were still high despite reduced range (r=.80 and r=.84 respectively). The correlation between Google and COCA was the lowest for the random pairs, r=.60, because 31 of the 40 random pairs had frequencies of 0 in COCA.

Place Table 1 here

The correlations between corpus frequencies and familiarity ratings were also high (both r=.88; see top panel of Table 1). To better understand how familiarity judgments relate to corpus frequencies, we graphed familiarity ratings as a function of COCA frequencies. As shown in Figure 1, a floor effect occurred for the random pairs. The dense cluster in the lower left side of graph occurred because 31 of the 40 random pairs were absent from COCA. Many random pairs had ratings near 0 because the majority of raters judged them non-sensical. The graph also shows that some of the legal pairs overlapped with the collocation frequency range, and considerable overlap existed in COCA frequency between the low and high collocations. It was thus decided, for the remaining analyses, to eliminate the low/high frequency division that had been determined using the HAL corpus, and to reclassify four collocations as merely legal combinations. These were the 4 that had the lowest frequency, and which met the criterion of being a legal constituent and lacking a strong idiomatic quality (the 4 word pairs were *mind bomb, small fuss, pay rate, sale ends*).

Place Figure 1 here

The relationship between familiarity ratings and Google frequencies (not shown) is broadly similar to the relationship between familiarity and COCA frequencies, as would be expected by the overall r=.93 between COCA and Google frequencies. A difference is that the use of Google as the frequency metric produced a scatter plot that, compared to Figure 1, is more extended for the random pairs. The top panel of Table 1 provides correlations separately for familiarity and the two frequency metrics, for each type of word pair. Familiarity ratings for the random pairs were not related to COCA frequency (r=-.02), but were weakly related (r=.31) to Google frequency. One of the differences between COCA and Google is that all the random pairs had an existence in Google -- and apparently not a random existence, ratings correlated with Google frequencies.

It is interesting that familiarity ratings for legal pairs were related to corpus frequencies, although moderately, and indeed, correlations for the legal pairs were similar in magnitude to the correlations for the collocations. The plot in Figure 1 and correlations in Table 1 thus support our first goal, which was to show frequency effects on behavioral responses outside of high frequency correlations. Familiarity ratings correlated with corpus frequencies not just for the cases that would be predicted to do so by most theories, i.e., well-known collocations, but also for adjective+noun pairs which are generally considered compositional (e.g., *early change, some cans*), and randomly combined word pairs.

The next section analyzes our second behavioral measure, perceptual identification of the word pairs.

2.3 Correlations between perceptual identification and frequency/familiarity

Participants' ability to identify the words (perceptual identification or PID task) correlated overall relatively strongly with COCA (r=.58) and Google corpus statistics (r=.61). Obviously because of restricted range, correlations for the word types analyzed separately were weaker, but importantly, for the legal pairs, PID correlated with COCA (r=.33) and Google frequencies (r=.34; see bottom panel in Table 1). For random pairs, the highest correlation (r=.39) was with Google frequencies. Lower r values for COCA and familiarity likely resulted from floor effects. As noted above, for familiarity, floor effects occur because raters had the option of circling "doesn't make sense." Future work could explore how sensitive raters are to variations in meaning and familiarity within the category of words which are ostensibly unrelated but which can occur in print adjacent to each other, as demonstrated by above 0 Google frequencies. The finding that probability of perceptual identification for these low frequency and 0-frequency items is related to Google frequencies (and more weakly, to COCA frequencies, see Figure 2) is thus particularly impressive.

To graphically depict these frequency effects, Figure 2 plots perceptual identification as a function of COCA log frequency. Word pairs within each category were split into low and high frequency categories for purposes of illustration. The mean of the log frequencies for both low and high are plotted with standard error shown with error bars. This shows that there was considerable overlap at the high end of the random and at low end of the legal pairs in both COCA frequencies and perceptual identification.

Place Figure 2 here

Many of the random pairs are so nonsensical that in Google they mainly occur separated by punctuation or graphical white space (e.g., *weep job* straddles a period and references the Biblical parable of Job). Some can be assimilated to an adjective noun construction, e.g., *belt trade, trick boy*, and these had higher Google frequencies and better perceptual identification. Future work will need to determine whether frequency effects for random pairs reduce to a difference between two types of random pairs: those that can and can't be readily assimilated to a grammatical pattern.

3 Frequent recitation of prayers creates 'tracks in the mind'

Above we tested the usage-based hypothesis and emergentist accounts by demonstrating frequency effects for word pairs across the frequency continuum, from high frequency collocations, to low frequency collocations, to merely legal combinations, to random word pairs. Another prediction of the usage-based model is that language users who have more experience with specific linguistic stimuli will have more efficient processing of those stimuli. Psycholinguists have not measured individual differences in stimulus expertise as routinely as have cognitive neuroscientists who have shown, for example, how expertise with specific objects influences brain organization (Bukach, Gauthier and Tarr 2006). One challenge is identifying speakers who reliable differ in their language experience. Language researchers have examined variation in exposure to language by comparing native vs. second language learners (e.g., Ellis and Simpson-Vlach 2009; Gilquin and Gries 2009). Studying second language learners is certainly a good way to identify groups with more vs. less usage, but many aspects of language use are altered for non-native speakers in addition to reduced usage. It would thus be ideal to find groups who reliably vary primarily in their usage of specific expressions. An example of prior work which did this in a compelling manner is Stadthagen-Gonzalez, Bowers, and Damian (2004), who used professional expertise to detangle frequency effects from age-of-acquisition effects, arguing that acquisition age would presumably be similar for a word like *cognition* for both chemists and psychologists, but only high frequency for the psychologists.

3.1 Individual differences in prayer habits

Our interest in collocations and routinized patterns suggested that phrases from religious rituals would have different patterns of use across groups with different prayer habits. Observant Orthodox Jews are required to recite three prayers every day. The linguistic sequences in these daily prayers would presumably be quite entrenched, compared to weekly and annual prayers. By comparing Orthodox Jews to secular Jews (and also directly inquiring about prayer recitation practices) one would have two groups with different usage patterns.

Studying Jewish prayers is a particularly fertile area because daily, weekly and annual prayers exist. We studied phrases from weekly and annual prayers, with the proviso that using such phrases is highly exploratory. Frequency may not be the most important factor for entrenchment because some prayers may have greater emotional resonances than other prayers. Weekly prayers recited on Saturdays are longer than daily prayers and occur with a different service. The annual prayers recited over the High Holy Days are further prolonged and the services carry a higher emotional charge than do services accompanying the daily and weekly prayers.

Given that one third to one-half of the Jewish population in Israel is secular, nonreligious Israelis could readily be recruited as a comparison group. Secular people generally do not recite the daily or weekly prayers, but many do attend the annual services during the High Holy Days.

3.2 Method

Participants self-identified as religious (N=32, 19 females and 13 males) or secular (N=19, 11 females and 8 males). Each participant completed a questionnaire detailing praying habits (frequency of praying and whether in private or at synagogue). We additionally measured participants' knowledge of Jewish prayer texts using a phrase completion test. This test consisted of 17 phrases taken from various Jewish prayers. For examle, the phrase *barux shem k'vod* ____ would be finished with *Malchuto L'olam Va'ed* (meaning of whole phrase: *Blessed is the name of his glorious kingdom for all eternity)*. Phrases that were left blank or were completely wrong received 0 points, one point was given for partial completion, and two points were given for perfect completion of the phrase.

Materials for the perceptual identification task were six types of phrases which were selected to have comparable semantic and syntactic complexity (see Table 2). Religious phrases were categorized according to frequency of recitation (daily, weekly, and annual). Nonreligious phrases were selected to be either common or rare. The common phrases were drawn from Israeli culture and included political slogans, names of famous TV shows, and popular songs. The rare phrases were selected from modern Hebrew literature and poetry. Google counts confirmed that the phrases in the rare group were substantially less common than phrases in the common group (mean log frequency of rare phrases = 1.6, common phrases = 4.6; $p<0.0001$). The sixth group was constructed out of words that appear separately from each other in Jewish prayers and do not form cohesive phrases when mixed. Each phrase group comprised eleven 2-word phrases and four 3- word phrases. The length in characters of phrases was similar across all categories (mean=12.1, std=0.55).

Place Table 2 Here

To avoid floor and ceiling effects in the perceptual identification task, exposure durations were set individually for each participant, based on performance in practice trials, with the average exposure duration for two-word phrases 71 ms and for three-word phrases, 90 ms. In order to obtain as much information as possible about participant's perception of the stimuli, we scored their accuracy on a 3 point scale for no words correctly reported, partial correct report of the phrase (at least one word correct), and complete report of the target phrase.

3.3 Results and Discussion

Compared to the secular group, religious participants were more accurate on the religious phrases and showed stronger frequency effects. This was confirmed via ANOVA with religious group and stimulus frequency as between-subject and within-subject predictors. Main effects were obtained in the expected directions for religious group, $F(1,47)=15.5$, $p<0.001$, and frequency $F(2,96)=33.7$, $p<0.0001$). The two groups did not differ in their accuracy for the common and rare secular phrases, $F < 1$, but both groups showed frequency effects, $F(1,48)=470.7$, $p<0.0001$ (see Figure 4). Additional exploratory analyses, including analyses of gender effects, phrase completion scores, and self-reported prayers habits, are reported separately (Berant, Caldwell-Harris and Edelman 2008).

Place Figure 4 here

As noted above, phrases from daily prayers had higher accuracy than phrases selected from weekly and annual prayers, and these frequency effects were stronger in religious participants. What had not been predicted is that non-religious participants were also affected by the frequency of phrases, but in an attenuated and different manner, as indicated statistically by the group X frequency interaction, $F(2,96)=5.5$, $p<0.01$. For secular participants, accuracy was highest for the daily and annual prayers, and lowest for the weekly prayers (illustrated in Figure 3). It is possible that the weekly phrases were unintentionally more difficult than the daily and annual phrases. The explanation we favor is that annual phrases were more entrenched than would be expected by a once-yearly recitation, because of their high emotional charge. This can explain the relatively good accuracy for the annual prayers shown by religious participants but is probably especially true for the secular participants. Secular participants may attend synagogue during the annual High Holy Days, an occasion that is memorable.

Place Figure 3 here

This study of perceptual identification of phrases from Jewish prayers directly supports the usage-based model of human language. Individuals who have greater experience with specific linguistic expressions had greater accuracy at reading the briefly displayed phrases, consistent with the predictions from ADIOS that frequently encountered sequences lay down 'tracks in the mind.'

4 General Discussion

4.1 New evidence for the pervasiveness of frequency effects

The analysis of word pair data demonstrated frequency effects not just for high frequency common word combinations, but for low frequency collocations, and for word pairs which are merely legal combinations (*some cans*). Frequency effects were found even among two word sequences that had been randomly put together and had zero frequency in COCA. This finding thus moves beyond the results of Arnon and Snider (2010) who found that response latencies varied continuously across a range of low and high frequency collocations (four-word sequences). However, the low frequency stimuli in Arnon and Snider's study had a minimum occurrence of 1 per million and extended to 9 per million. Our legal pairs averaged .35 per million, and our random pairs had an average frequency of .003 per million, occurring on average only 1.3 times in the 410 million word COCA. Models which propose that statistics are maintained or exemplars stored only for sequences with some minimum frequency will find it difficult to account for these frequency effects.

The goal in our study of processing of Jewish prayer phrases was to determine if individual differences in verbal expressions reliably led to processing differences, as predicted by the usage-based hypothesis and emergentist models like ADIOS (Solan et al. 2005). Religious Jews had better identification of phrases from daily prayers than for weekly or annual prayers. Compared to religious Jews, secular Jews had overall poorer identification of the religious phrases and showed only weak frequency effects. This is strong support for the usage-based hypothesis. We hope these results encourage other researchers to undertake individual differences research.

4.2 Why are language users sensitive to the frequency of word sequences?

Strong and diverse effects of frequency were found across these two studies. Why does the brain keep track of these statistics? Are frequency statistics useful for a real task in comprehension or in production, or are they a by-product of something else?

Humans plausibly have statistical information about the frequencies of word combinations because they store exemplars. Storing exemplars aids both acquisition and processing. As experience with the ADIOS algorithm shows, statistics about word sequences are essential to the ability to infer grammatical structure (Solan et al. 2005). Moreover, processing benefits are likely to result when people can rely on stored constructions (e.g., Lewis and Vasishth 2005), especially for highly proficient language users. Storing common phrases allows listeners to anticipate upcoming words, allowing the use of top-down expectations to clean up a noisy speech signal or infer the completion of a sentence when only the first part has been received. More generally, storing frequency-weighted exemplars helps in predicting the world, a crucial information processing strategy that language processing shares with much of the rest of cognition (Edelman 2010).

4.3 How are word pairs stored?

One can imagine a phrasal level of representation in which top-down activation from the phrasal level explains why collocation status strongly influenced accuracy of word identification. This could be analogous to the word superiority effect, in which orthographic regularities in an individual word facilitate letter recognition (e.g., McClelland and Rumelhart 1981). Top-down and bottom-up interactive-activation and competition could explain how collocation frequency aids recognition of words which are rapidly sequentially displayed. In the case of word pairs, statistical regularities between word pairs could boost recognition of the individual words.

Algorithms like ADIOS (Solan et al. 2005) augment the interactive activation account by specifying a computational procedure whereby an initially flat representation of utterances as strings of words becomes hierarchical with experience, with collocations being assigned their own units.

Many psycholinguists propose that regularity and entrenchment in language is a matter of degree (McClelland 2002; Seidenberg and McClelland 1989). A continuum of unitization may exist, with fully fused word pairs like *blackboard* at the one extreme end, *middle name* and *last chance* occupying an intermediate position and rare and novel combinations at the non-entrenched end of the continuum (Harris 1998; Wray 2002). In computational modeling, gradedness is a more revolutionary concept, and indeed many models implement unitization. The original interactive activation model (McClelland and Rumelhart 1981) assumed unitization, and ADIOS also currently assumes unitization. This may change as modelers grapple with and surmount the computational challenges of implementing gradedness.

4.4 Converging corpora and behavioral data

In the current paper we answered the plea of Arppe et al. (2010) and Gilquin and Gries (2009) for converging data, including relating different frequency measures to each other. The study of word pairs compared two behavioral measures (familiarity and perceptual identification) and employed two frequency corpora, COCA and Google. Very high correlations were obtained between COCA and Google. It is noteworthy that the correlation between COCA and Google was high even for sequences which are not typically considered multiword utterances, the legal pairs, (r=.84), and was at least moderate for random pairs (r=.60). This indicates that these word pairs have reliable frequency statistics in the world of printed materials. The correlations between corpora frequencies and familiarity ratings were also strong, especially when calculated across the frequency range (e.g., r=.87 for all 160 items). This extends to two-word pairs the observation of Balota, Pilotti and Cortese (2001) that native speakers can reliably estimate words' relative frequencies.

Why is familiarity a better fit to corpora frequencies than perceptual identification? Corpora frequencies and subjective familiarity performed similarly in predicting the perceptual data from the fast pairs paradigm described above. However, perceptual identification had smaller correlations with objective frequency than did subjective familiarity. We speculate that familiarity ratings emerge from a process that normally has sufficient time to settle into a stable state, while perceptual phenomena are influenced by more random variables, such as attention and momentary physiological factors.

5 Open questions and future work

5.1 Merely legal and barely legal: Why are frequency effects obtained?

Future work on processing of merely legal word pairs such as *her list* and barely legal items such as *edit center* can determine why frequency effects occur for these items. The goal of the current analysis was to determine the extent of frequency effects, and the data set was not constructed to test a range of causal factors. Items in the random group such as *butter ace* and *cast bark* received below zero familiarity ratings, meaning more than half of raters judged them to make no sense. These items were absent in COCA, had low frequency in Google, and also had poor recognition on the perceptual identification task. Consider random pairs with above zero familiarity ratings such as *work use* and *edit center*. These appear easily assimalable to an adjective-noun category, and also had higher corpus frequencies and perceptual identification scores than other random pairs. Being assimilable to an adjective-noun category could cause raters to avoid using the "doesn't make sense" label. But why are corpus frequencies higher for these items? Systematic analysis of contexts across a large set of items is required to determine if these mostly appear as adjective+noun pairs, of if they are appearing in other syntactic contexts, e.g., from Google, *How to edit Center Ring*.

Does merely appearing in contiguous order influence mental entrenchment, even if the word pair is not a constituent? For example, a random pair like *city away* is not assimilable to any syntactic constituent, but gains its occurrences in corpora in sequences like *Flood of Complaints Washes Tent City Away* (from Google). Are raters reliably sensitive to the semantic features in *city* such as geographical location which match features in the adverb *away*? We speculate that that off-line rating may be more sensitive to semantic feature match, while the ability to perceive a phrase under degraded conditions is more sensitive to prior experience of contiguity.

5.2 Is constituency more important than mere contiguity?

Computational models of language acquisition like ADIOS assume language learners initially collect statistics about word and phrase co-occurrences, in order to infer grammatical constructions via their distributional regularities. What remains unclear is what types of exemplars and frequencies are retained once typical phrase boundaries are learned. Do speakers shift to partly or fully collecting statistics that respect phrase boundaries?

5.3 What is the role of semantic meaningfulness and the emotional charge of expressions on how deeply entrenched linguistic sequences can become?

Corpus frequencies will be only one factor influencing entrenchment and mental representation, as has been discussed by many authors (Snider and Arnon this volume; Caldwell-Harris and Morris 2009; Gilquin and Gries 2009). Other factors may be semantic coherence, grammatical constituency, and emotional resonance. Ellis and Simpson-Vlach (2009) reported that mutual information (a statistical metric, Barlow 1990) was a better predictor of native-speakers' judgments of multiword expressions. For polysemous words, concrete senses (e.g., the "hand over" sense of *give*) are the most salient to raters, but phraseological uses (*give me a smile*) are most frequent in the corpora (Arppe et al 2009).

Langacker (1987) has noted that accruement of semantics beyond compositionality is a pervasive feature of language and occurs at every unit of analysis, from the morphological level (e.g., the common referent of *stapler* is not an object that staples but a specific instrument with recognizable shape) to the sentential level (e.g., *She felt the baby kick* typically refers to a pregnant woman feeling the kick of a fetus). It does intuitively seem that additional semantic coherence accrues to many commonly occurring multi-word utterances, a phenomenon discussed by Arnon and Snider (2010; see also Snider and Arnon, this volume).

When native speakers rate the familiarity of common word combinations, they may be influenced by emotional resonances of the overall meaning of the words, leading them to rate emotional phrases such as *child abuse* and *caring words* as more familiar than objectively more frequent phrases such as rather than. The current corpus of 160 word pairs contained collocations such as *death bed, face value* and *upper hand* which contain an idiomatic quality. That is, the specific meaning of the combination is not fully predictable from component words, or at least the word pair identifies referents beyond what would be inferred on an adjective+noun analysis (e.g., *gold medal* is a specific type of award, not just a medal that is gold; *black hole* is not just a hole that is black). While semantic coherence was not a requirement for selection in the current corpus, most of the collocations in the word pair study are either idiomatic or have specific culturally acknowledged referents, while most of the legal pairs lacked this: compare the legal pair *blue wall* to the collocation *green light*. It is possible that this extra referentiality of the collocations and/or extra emotionality provides them with a recognition boost (see higher slope for collocations in Figure 2).

The data from the Jewish prayers study also contained tantalizing hints that multiple factors influence entrenchment, in addition to frequency. Secular Jews recognized annual phrases with the same accuracy as daily phrases, with weekly prayers having lower accuracy. We speculate that daily prayers benefit from being commonly known because they are supposed to be recited daily, and secular Jews may are aware of them as part of being familiar with Jewish cultural knowledge. But annual prayers benefit from the emotional charge of the High Holy Days. Future work on within-speaker variation in mental entrenchment can investigate whether "tracks in the mind" are mainly influenced by amount of exposure vs. personal emotional response to the stimuli.

Acknowledgments

We thank Mark Davies for providing us with the COCA frequencies for the 160 word pairs in the Appendix.

Bibliography

Arnon, Inbal and Neal Snider 2010 More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language, 62,* 67-82.

Arppe, Antti, Gaetanelle Gilquin, Dylan Glynn, Martin Hilpet, and Arne Zeschel. 2010 Cognitive corpus linguistics: Five points of debate on current theory and methodology. *Corpora 5,* 1-2

Balota, David A., Maura Pilotti and Michael J. Cortese 2001 Subjective frequency estimates for 2,938 monosyllabic words. *Memory and Cognition 29,* 639-647.

Bannard, Colin and Danielle Matthews 2008 Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science, 19,* 241-248.

Barlow, H.B. 1990 Conditions for versatile learning, Helmholtz's unconscious inference, and the task of perception. *Vision Research 30* 1561-1571.

Beckner, Clay, and Joan Bybee. 2009. A usage-based account of constituency and reanalysis. *Language Learning 59,* 27-46.

Berant, Jonathan B., Shimon Edelman, and Catherine. L. Caldwell-Harris 2008 Tracks in the mind: Differential entrenchment of common and rare liturgical and everyday multiword phrases in religious and secular hebrew speakers. Presented at the *Annual Meeting of the Cognitive Science Society*, Washington, D.C., July 2008.

Bybee, Joan, and Paul Hopper (eds.) 2001 *Frequency and the Emergence of Linguistic Structure*. John Benjamins Publishing Company.

Bukach, C. M., Isabelle Gauthier, and Michael J. Tarr 2006 Beyond faces and modularity: the power of an expertise framework. *Trends in Cognitive Sciences 10,*159-166.

Caldwell-Harris, Catherine L. and Alison L. Morris 2008 Fast Pairs: A visual word recognition paradigm for measuring entrenchment, top-down effects, and subjective phenomenology. *Journal of Consciousness and Cognition* 17, 1063-1081.

Caldwell-Harris, Catherine L. 2008 Language research needs an "emotion revolution" AND distributed models of the lexicon. *Bilingualism: Language and Cognition* 11, 169-171.

Carr, Thomas H. 1986 Perceiving visual language. In K. R. Boff, L. Kaufman and J. P. Thomas (eds.), *Handbook of perception and human performance*. New York: Wiley.

Davies, Mark. 2010 Corpus of Contemporary American English COCA. Available from http://www.americancorpus.org/

Divjak, D.S. 2008 On (in)frequency and (un)acceptability. In Barbara Lewandowska-Tomaszczyk (ed.), *Corpus Linguistics, Computer Tools and Applications - state of the art* (pp 213-233). Frankfurt: Peter Lang..

Edelman, Shimon 2010 On look-ahead in language: navigating a multitude of familiar paths. In M. Bar (ed.), *Prediction in the Brain*. Oxford, UK: Oxford University Press

Eder, Andreas B., Bernhard Hommel, and Jan De Houwer 2007 How distinctive is affective processing? On the implications of using cognitive paradigms to study affect and emotion. *Cognition and Emotion 21*. 1137-1154

Ellis, Nick C. and Diane Larsen-Freeman 2009 *Language as a Complex Adaptive System*. *Wiley*-Blackwell

Ellis, Nick C. and Rita Simpson-Vlach 2009 Formulaic language in native speakers: Triangulating psycholinguistics, corpus linguistics, and education. *Corpus Linguistics and Linguistic Theory 5*, 61-78.

Ellis, Nick. C., and Diane Larsen-Freeman (eds.)2006 Language emergence: Implications for Applied Linguistics [Special issue]. *Applied Linguistics 27*.

Elman, Jeffrey L. 1995 Language as a dynamical system. In Robert F. Port & T. van Gelder (Eds.), *Mind as Motion: Explorations in the Dynamics of Cognition*. Cambridge, MA: MIT Press. Pp. 195-223.

Gibbs, Raymond W. and Nandini P. Nayak 1989 Psycholinguistic studies on the syntactic behaviour of idioms. *Cognitive Psychology 21,* 100-138

Gilquin, Gaetanelle and Stefan Th. Gries. 2009 Corpora and experimental methods: a state-of-the-art review. *Corpus Linguistics and Linguistic Theory 5,* 1-26.

Harris, Catherine L. 1998 Psycholinguistic studies of entrenchment. In J. Koenig (Ed.), *Conceptual Structures, Language and Discourse, Vol 2*. Berkeley, CA: CSLI.

Jackendoff, Ray1995 The boundaries of the lexicon. In M. Everaert, E. J. van der Linden, A. Schenk and R. Schreuder (Eds.), *Idioms: Structural and psychological perspectives*. Hillsdale, NJ: Erlbaum.

Langacker, Ronald W. 1987 *Foundations of cognitive grammar. Vol. I: Theoretical prerequisites*. Stanford, CA: Stanford University Press.

Lewis, Richard L. and Shravan Vasishth 2005 An activation-based model of sentence processing as skilled memory. *Cognitive Science 29,* 375-419.

Lund, Kevin and Carl Burgess 1996 Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, and Computers* 28 203–208.

MacWhinney, Brian Ed., 1999 *The emergence of language*. Mahwah, NJ: Erlbaum.
McClelland, James L. and Kathryn Patterson 2002 `Words Or Rules' cannot exploit the regularity in exceptions: Reply to Pinker and Ullman. *Trends in Cognitive Sciences 11,* 464-465.
McClelland, James L., Mark S. Seidenberg, and Kathryn Patterson 1996 Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review 103,* 56–115.
McClelland, James L. and Davd E, Rumelhart 1981 An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review* 88, 375–407.
Niedeggen, Michael, Martin Heil and Catherine L. Harris 2006 "Winner take all" competition among real and illusory words. *NeuroReport* 3, 493-497.
Pennebaker, James W., Martha E. Francis and Roger J Boot 2001 *Linguistic Inquiry and Word Count: A computerized text analysis program*. Mahwah, NJ: Erlbaum.
Pinker, Steven 1999 *Words and rules: the ingredients of language*. New York: Basic Books.
Pustejovsky, James 1995 *The Generative Lexicon*. Cambridge, MA: MIT Press.
Ratcliff, Roger, & Gail McKoon 1997 A counter model for implicit priming in perceptual word identification. *Psychological Review 104*, 319-343.
Stadthagen-Gonzalez, Hans, Jeffrey S. Bowers, and Markus F. Damian. 2004 Age-of-acquisition effects in visual word recognition: evidence from expert vocabularies. *Cognition 93,* B11-B26.
Swinney, David, and Anne Cutler 1979 The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior 18,* 523-534.
Solan, Zach, David Horn, Eytan Ruppin, and Shimon Edelman 2005 Unsupervised learning of natural languages. *Proceedings of the National Academy of Sciences* 102, 11629-11634.
Toglia, Michael, P. and William F. Battig 1978 *Handbook of semantic word norms*. Hillsdale, NJ: Erlbaum.
Tomasello, Michael 2003 *Constructing a language: a usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
Waterfall, Heidi R., B. Sandbank, Luca Onnis, and Shimon Edelman 2010 An empirical generative framework for computational modeling of language acquisition. *Journal of Child Language*.
Wray, Alison 2002 *Formulaic language and the lexicon*. Cambridge, UK: Cambridge University Press.

Table 1. Correlations

Correlations Between the Two Corpus Frequencies, COCA and Google

	N	r	Statistical significance
All items	160	**0.93**	$p < .0001$
Collocations	76	**0.80**	$p < .0001$
Legal Pairs	44	**0.84**	$p < .0001$
Random Pairs	40	**0.60**	$p < .0001$

Correlations Between Familiarity Ratings and Corpus Frequencies

	N	COCA	Google	Stat. significance
All items	160	**0.87**	**0.88**	both $p < .001$
Collocations	76	**0.47**	**0.46**	both $p < .001$
Legal Pairs	44	**0.44**	**0.52**	both $p < .005$
Random Pairs	40	0.04	**0.31**	Google, $p < .05$

Correlations Between Perceptual Identification and Familiarity/Frequency

	N	COCA	Google	Fam.	Stat. significance
All items	160	**0.59**	**0.61**	**0.61**	all $p < .02$
Collocations	76	**0.22**	0.15	**0.34**	Fam, $p < .005$; COCA, $p=.057$
Legal Pairs	44	**0.32**	**0.34**	0.22	both $p < .03$
Random Pairs	40	0.23	**0.39**	0.29	Google, $p < .02$

Table notes. Bold r values are statistically significant

Table 2: Example Stimuli for Jewish Prayers Study

Phrase Type	Example and translation	
Daily prayers	*morid hatal*	he who makes the dew drop down
Weekly prayers	*nafshi yeshovev*	will exhilarate my spirit
Annual prayers	*bnei maron*	sheep and goats (archaic)
Common phrases	*shalom xaver*	good bye, friend
Rare phrases	*divrey rahav*	words of arrogance
Random	*zore'a ha'amim*	sower of nations

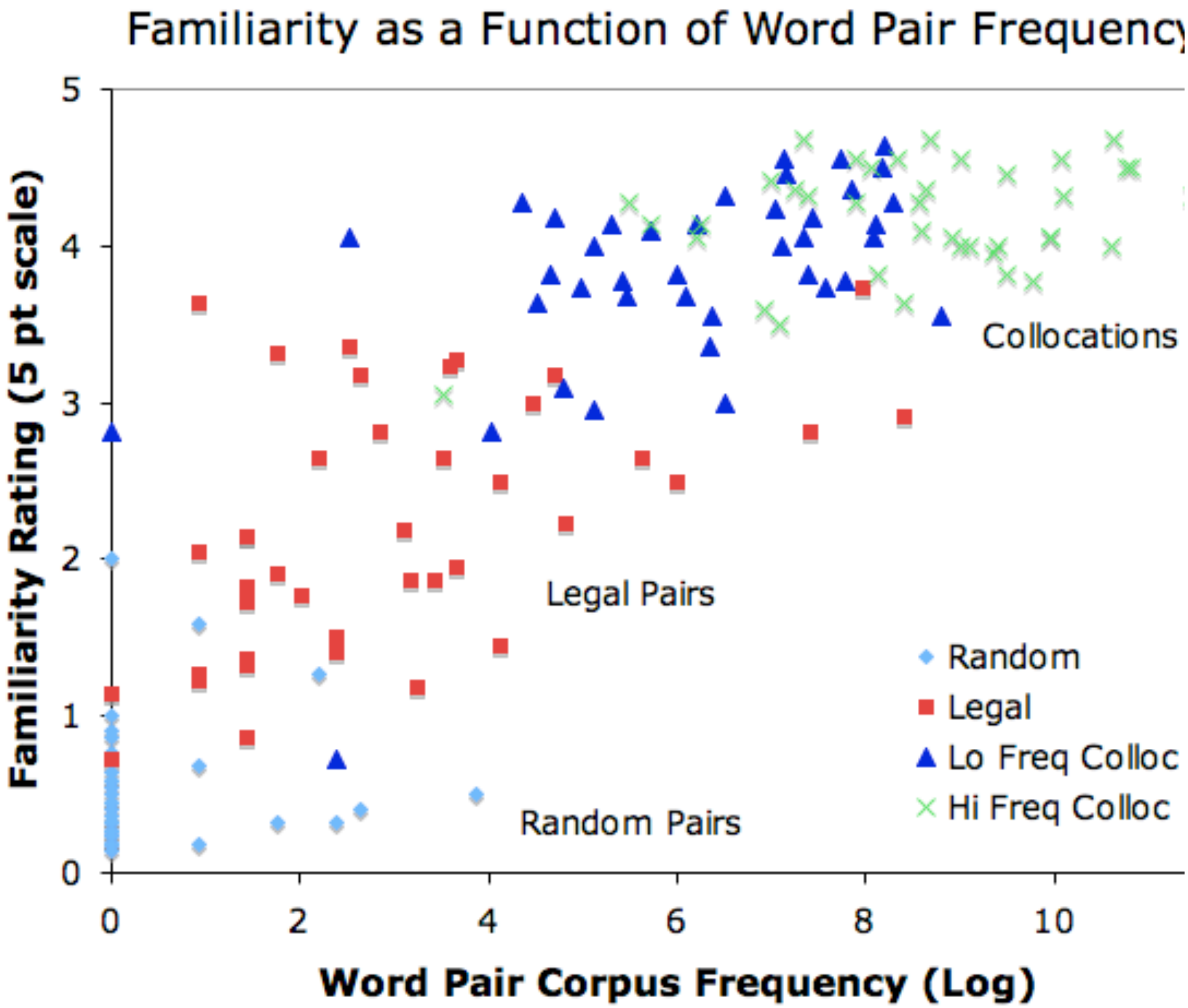

Figure 1. The relationship between students' familiarity ratings of the 160 word pairs, and the pairs' frequencies in COCA (Davis 2010), with collocations grouped according to low and high frequency collocations as defined by the HAL corpus (Lund and Burgess 1996).

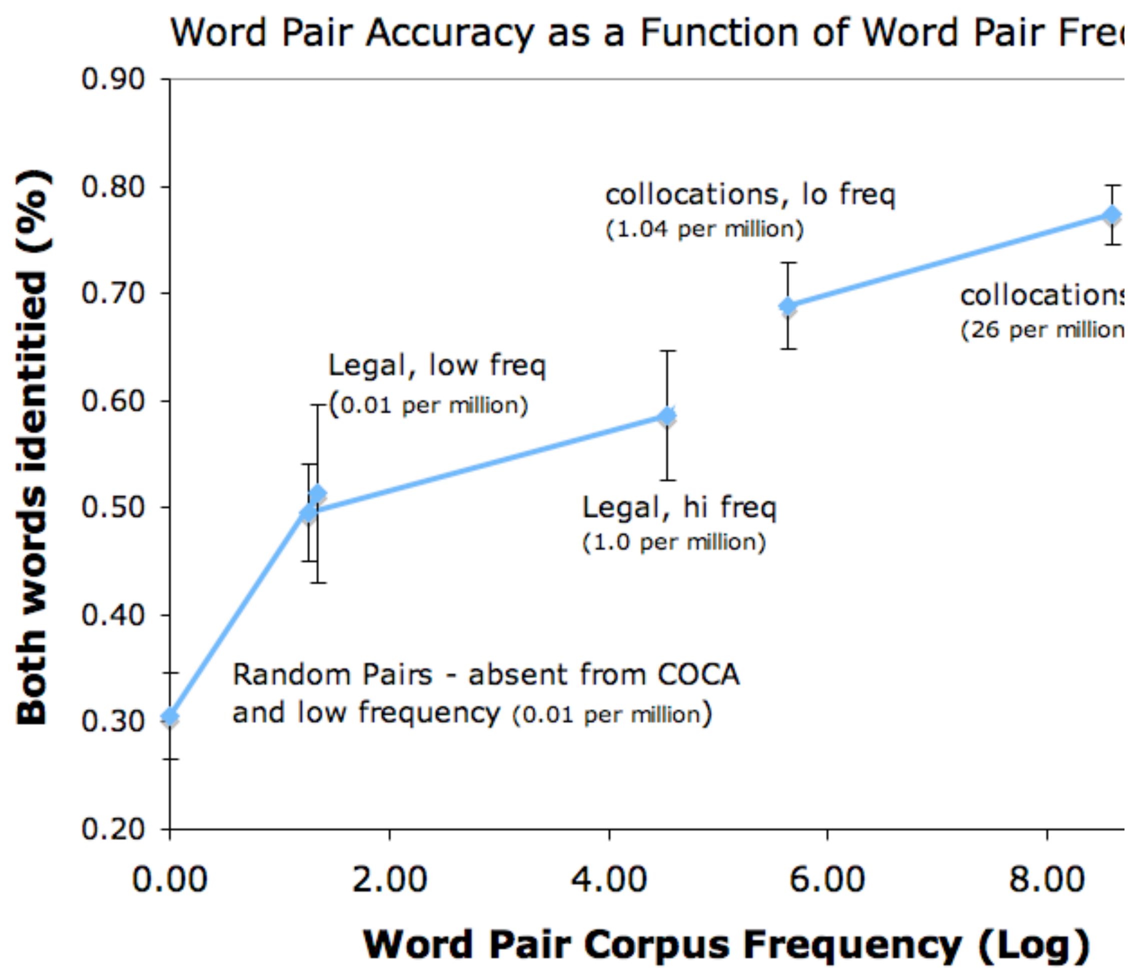

Figure 2. Perceptual identification plotted as a function of COCA frequencies, with low and high frequency categories defined by COA frequencies.

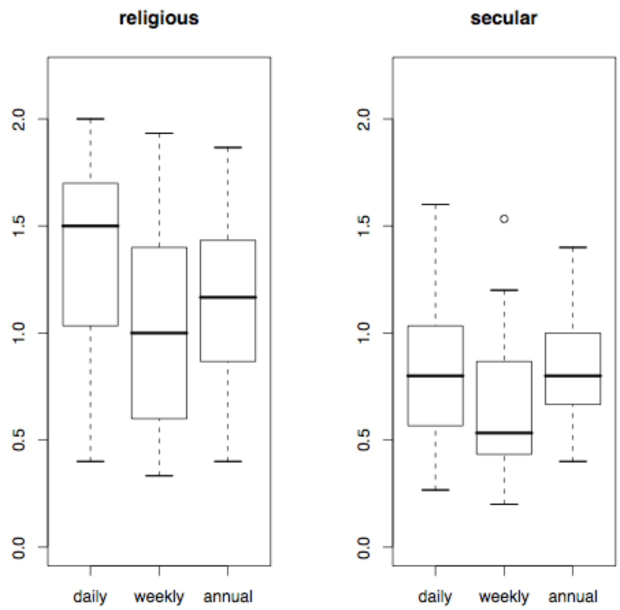

Figure 3. Mean correctness for religious participants compared to secular participants for phases taken from daily, weekly and annual prayers. Compared to secular participants, religious participants identified more words from the phrases, and showed stronger frequency effects.

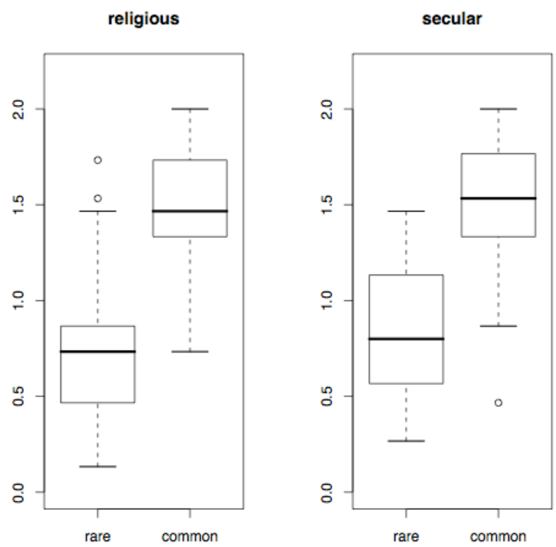

Figure 4. Correctness for nonreligious phrases did not vary according to individual religiousness; but both groups of participants more accurately identified the more frequent phrases.

Appendix: Data analyzed in word pair frequency study

Item		COCA(raw)	Google(Log)	Familiarity	Perceptual Identification
Random Pairs					
weep	job	0	3.85	0.32	0.29
butter	ace	0	4.06	0.5	0
comedy	span	0	4.08	0.91	0.29
pigs	troop	0	4.32	0.41	0
victim	cheese	0	4.62	0.14	0.25
course	hoop	0	5.2	0.64	0.29
basis	coast	0	5.32	0.27	0
cast	bark	0	5.47	0.18	0.415
blood	plane	0	5.87	0.27	0.165
pants	cloud	0	5.87	0.23	0.29
look	fever	0	5.98	0.32	0.29
desk	marks	0	6.14	2	0.25
cash	tone	0	6.19	0.45	0.415
belt	trade	0	6.29	0.77	0.875
taxi	tie	0	6.47	0.27	0.415
blast	brick	0	6.54	0.59	0.46
heart	root	0	6.58	0.18	0.25
smoke	bone	0	6.79	0.55	0.29
eyes	trees	0	7.01	0.32	0.415
art	beard	0	7.16	0.36	0
puppy	hill	0	7.22	0.45	0.25
anchor	stream	0	7.41	0.73	0.125
mass	floor	0	7.67	1	0.5
stroke	break	0	8	0.73	0.25
school	belly	0	8.48	0.18	0.25
hey	wing	0	8.68	0.18	0.25
trick	boy	0	9.38	0.59	1
dime	finger	0	9.68	0.18	0.29
pin	since	0	10.15	0.23	0.335
edit	center	0	10.52	1.14	0
days	group	0	11.66	0.86	0.58
wife	board	1	6.55	0.18	0.415
golf	where	1	11.21	0.68	0.25
home	leg	1	11.41	1.27	0.835
work	use	1	13.77	1.59	0.5
while	base	3	10.95	0.32	0.415
city	away	5	11.59	1.27	1
war	say	6	12.55	0.32	0.455
cold	off	8	10.81	0.41	0.5
system	never	28	12.9	0.5	0.25

Legal Pairs

silly	trail	0	6.64	0.73	0.46
small	fuss	0	7.66	2.82	0.75
happy	name	0	10.09	1.14	0.33
tough	sort	1	6.35	2.05	0.165
proper	widow	1	8.16	1.23	0
dead	bride	1	9.91	1.27	0.29
caring	words	1	10.47	3.64	0.415
social	bunch	2	6.49	1.82	0.75
lousy	law	2	7.44	2.14	0.125
cruel	cost	2	8.52	1.32	0.46
famous	angel	2	9.5	1.36	0.75
modern	barn	2	9.57	0.86	0.25
simple	trend	2	10.59	2.14	0.71
right	unit	2	11.32	1.73	0.415
huge	church	3	10.99	3.32	0.5
left	step	3	11.17	1.91	0.415
early	change	4	11.13	1.77	1
true	people	5	13.11	2.64	0
mind	bomb	6	11.46	0.73	0.75
real	skin	6	11.86	1.41	0.58
same	run	6	12.69	1.5	0.625
entire	survey	7	12.28	3.36	0.375
sale	ends	7	14.59	4.05	0.71
size	three	8	12.29	3.18	0.54
public	attack	10	11.01	2.82	0.585
some	cans	13	11.46	2.18	0.75
gray	eye	14	10.41	1.86	0.54
empty	world	15	11.87	1.18	0.705
such	space	18	13.91	1.86	0.415
lost	girl	20	13.01	2.64	0.75
pay	rate	20	15	3.05	0.71
green	skirt	21	11.47	3.23	0.54
best	woman	23	13.04	1.95	0.415
new	table	23	14.43	3.27	1
open	spot	37	12.22	2.5	0.29
night	man	37	12.77	1.45	0.165
major	case	52	12.83	3	0.5
good	race	66	13.56	3.18	0.375
blue	wall	73	12.42	2.23	0.54
little	food	165	13.63	2.64	0.71
her	list	245	13.88	2.5	0.5
each	state	1001	17.21	2.82	0.585
two	ways	1726	17.9	3.73	0.71
both	men	2694	16.77	2.91	1

Collocations

cents	worth	33	14.21	2.82	0.835

full	refund	46	16.46	4.27	0.415
local	bus	54	15.71	3.64	0.415
killer	bees	62	13.18	3.82	0.415
death	bed	66	13.9	4.18	0.585
sole	reason	72	14.03	3.09	0.455
hot	wheels	88	15.8	3.73	1
eight	ball	99	15.16	4	1
any	clues	101	14.19	2.95	0.25
litter	box	121	14.63	4.14	0.875
book	sales	136	16.11	3.77	0.5
high	esteem	140	14.06	3.68	1
money	order	147	18.03	4.27	1
pet	store	182	15.51	4.09	0.71
own	risk	184	17.64	4.14	0.75
mere	fact	242	15	3.82	0.75
safe	bet	266	14.55	3.68	0.585
blind	date	294	15.4	4.14	0.71
junk	mail	295	16.69	4.05	0.875
fan	club	310	17.07	4.14	1
wonders	why	339	14.87	3.36	0.415
die	hard	355	16.5	3.55	0.83
next	phase	407	16.24	3	0.125
top	secret	408	16.67	4.32	0.25
must	admit	613	16.56	3.59	0.75
zip	code	658	19.66	4.41	0.71
upper	hand	688	15.34	4.23	0.75
low	level	723	17.92	3.5	0.58
fair	trial	735	15.38	4	1
face	value	751	16.76	4.55	0.46
rush	hour	763	13.36	4.45	0.835
phone	lines	843	15.99	4.36	0.875
feel	free	927	19.17	4.68	0.335
peace	corps	928	16.49	4.05	0.83
deep	inside	973	15.98	3.82	0.835
screwed	up	978	16.37	4.32	0.75
machine	gun	1017	15.96	4.18	0.71
broad	range	1160	18.45	3.73	1
talk	shows	1360	16.39	4.55	0.875
focal	point	1451	17.33	3.77	0.705
back	yard	1533	16.33	4.36	0.75
keep	track	1632	18.1	4.27	0.71
brown	hair	1635	15.94	4.55	0.705
vice	versa	1895	18.19	4.5	0.835
news	media	1943	18.13	4.05	0.58
gold	medal	1983	16.77	4.14	0.75
black	hole	2048	17.26	3.82	0.705
front	page	2126	19.76	4.5	0.835
child	abuse	2156	17.21	4.64	1

locker	room	2424	16.38	4.27	0.75
post	office	2486	18.16	4.55	1
suffer	from	2704	17.82	3.64	0.75
get	married	3168	16.66	4.27	0.71
based	upon	3248	18.77	4.09	0.705
long	term	3400	20.15	4.36	0.75
very	nice	3536	18.25	4.68	0.875
natural	gas	3983	18.24	3.55	0.335
all	sorts	4459	18.01	4.05	0.665
credit	card	4932	19.95	4.55	0.75
few	hours	4938	17.81	4	1
shut	down	5403	18.05	4	1
much	better	6968	18.68	3.95	0.75
sounds	like	7340	18.51	4	1
months	ago	7962	18.48	3.82	0.585
parking	lot	8017	17.71	4.45	1
great	deal	10644	18.74	3.77	0.875
figure	out	12705	18.88	4.05	0.415
worry	about	12795	18.52	4.05	0.71
too	many	14153	19.1	4.55	1
middle	east	14780	20.13	4.32	0.585
five	years	24529	19.44	4	0.75
every	day	24947	19.62	4.68	1
health	care	28623	20.52	4.5	0.835
last	week	30436	19.57	4.5	0.5
rather	than	57700	20.39	4.32	0.665
thank	you	77530	20.45	4.77	1