# The psycholinguistic reality of collocation and semantic prosody (1)

## Lexical access

Nick C. Ellis, Eric Frey and Isaac Jalkanen
University of Michigan

Our research investigates the psycholinguistic reality in language users of the phenomena of collocation and semantic prosody shown by corpus linguistics to be pervasive in language texts. This report concerns the earliest stages of word recognition and lexical access. It uses a lexical decision task to assess whether these processes are sensitive to particular collocations and to the generalizations of semantic prosody/association. The results demonstrate that native speakers preferentially process frequent verb-argument and booster/maximizer-adjective collocations. But the same paradigm that so readily shows sensitivity to particular collocations fails to demonstrate generalization. While memory for particular lexical associations affords fluent lexical access, there are no top-down semantic generalizations upon this level of processing. Our subsequent research shows semantic access to be the earliest cognitive locus of semantic association.

> You shall know a word by the company it keeps. (Firth 1957a)

## 1. Introduction

Fifty years on, corpus linguistic analyses of large collections of text have persuasively confirmed that natural language makes considerable use of recurrent patterns of words and larger constructions. Lexical context is crucial to knowledge of word meaning and grammatical role. One type of pattern is *collocation*, described by Firth as the characterization of a word from the words that typically co-occur with it. Sinclair[1] summarized the results of corpus investigations of such distributional

[1] Like very many in our field I have been deeply affected by John Sinclair and his work and I mourn our recent loss of the man. As I came to his work in 1993, he provided me, a psycholin-

regularities in the *Principle of Idiom:* "a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analyzable into segments" (Sinclair 1991: 110), and suggested that for normal texts, the first mode of analysis to be applied is the idiom principle, as most of text is interpretable by this principle. Kjellmer reached a similar conclusion: "In all kinds of texts, collocations are indispensable elements with which our utterances are very largely made" (Kjellmer 1987: 140). Erman and Warren (2000) estimate that about half of fluent native text is constructed according to the idiom principle. Comparisons of written and spoken corpora suggest that collocations are even more frequent in spoken language (Biber et al. 1999; Brazil 1995; Leech 2000). Collocations are patterns of preferred co-occurrence of particular words, like *blazing row* and *heated dispute* (but not *heated row* or *blazing dispute*). Another type of pattern is more abstract – the schemata that can be identified from the generalization across collocations. *Semantic prosody* refers to the general tendency of certain words to co-occur with either negative or positive expressions, "the consistent aura of meaning with which a form is imbued by its collocates" (Louw 1993: 157). A famous example, by Sinclair, is *set in*, which has a negative prosody: *rot* is a prime exemplar for what is going to set in. *Cause* (something causes an accident/catastrophe/other negative event), and *happen* (things go along smoothly, then "something happens", shit happens) similarly have a negative semantic prosody. These patterns come from usage – there are no defining aspects of the meaning of *cause* or *happen* which entails that they will take negative rather than positive objects. Hoey (2005; this volume) refers to such generalizations when a word or word sequence is associated in the mind of a language user with a semantic set or class as *semantic association*. Thus analyses of language *texts* demonstrate how lexis, grammar, meaning and usage are inseparable (Granger and Meunier 2008; Hunston and Francis 2000; Sinclair 1991, 2004).

Such observations of textual corpora naturally provoked linguists to make inferences about language *users* and about the cognitive processes of meaning, speech production and comprehension. The statement of the Principle of Idiom is a good example. Here are several others:

> Meaning by collocation is an abstraction at the syntagmatic level and is not directly concerned with the conceptual or idea approach to the meaning of words. One of the meanings of *night* is its collocability with *dark*… (Firth 1957b: 196)

> In the store of familiar collocations there are expressions for a wide range of familiar concepts and speech acts, and the speaker is able to retrieve these as

guist, with a theory of language that made sense and meshed, when other alternatives seemed distant and jarring.

> wholes or as automatic chains from the long-term memory; by doing this he minimizes the amount of clause-internal encoding work to be done and frees himself to attend to other tasks in talk-exchange, including the planning of larger units of discourse. (Pawley and Syder 1983: 192)

> for a great deal of the time anyway, language production consists of piecing together the ready-made units appropriate for a particular situation and ... comprehension relies on knowing which of these patterns to predict in these situations. (Nattinger 1980: 341)

> Suppose that, instead of shaping discourse according to rules, one really pulls old language from memory (particularly old language, with all its words in and everything), and then reshapes it to the current context: "context shaping", as Bateson puts it, "is just another term for grammar". (Becker 1983: 218)

> Every word is primed for use in discourse as a result of the cumulative effects of an individual's encounters with the word. If one of the effects of the initial priming is that regular word sequences are constructed, these are also in turn primed... The(se) are claims about the way language is acquired and used in specific situations. (Hoey 2005: 13)

> Corpus-based analysis can throw light on the nature and extent of collocational bonding between words... In addition, data of the kind considered here can reveal something of the cognitive processes which lie behind language learning and use, and which enable us to become fluent language users, and it is these insights which can be among the most satisfying of all. (Kennedy 2003: 485)

But however appealing these statements, they go beyond the data. While there is no denying that texts have been produced by language users, and thus must somehow reflect their thinking, corpus analyses say nothing about the cognitive loci of sensitivity of language learners and fluent users to these patterns of co-occurrence. The analysis of whether word recognition and lexical access, semantic activation, and the processes of production of speech and writing are sensitive to collocations, formulas, and the more abstract schemata potentially derivable from them, is an empirical matter, one that falls in a different domain of investigation, that of psycholinguistics.

Psycholinguistic research broadly confirms language users' sensitivity to various distributional aspects of language (Ellis 2002a, 2002b):

> Psycholinguistics is the testament of rational language processing and the usage model. The words that we are likely to hear next, their most likely senses, the linguistic constructions we are most likely to utter next, the syllables we are likely to hear next, the graphemes we are likely to read next, and the rest of what is coming next across all levels of language representation, are made more readily available

> to us by our language processing systems. Not only do we know the constructions that are most likely to be of overall relevance (i.e. first-order probabilities of occurrence), but we also predict the ones that are going to pertain in any particular context (sequential dependencies), and the particular interpretations of cues that are most likely to be correct (contingency statistics). These predictions are usually rational and normative in that they accurately represent the statistical covariation between events. In these ways, language learners are intuitive statisticians; they acquire knowledge of the contingency relationships of one-way dependencies and they combine information from multiple cues. (Ellis 2006: 7–8)

But psycholinguistic research also identifies a wide variety of largely separable processes of language cognition (Altman 1997; Gernsbacher 1994), and it demonstrates that these are *differentially* affected by factors such as type and token frequency, phonological, orthographic, morphosyntactic, grammatical and pragmatic consistency of pattern, cohort density and consistency, word class, imageability, age of acquisition, etc. (Ellis 2002a; Harley 1995; Levelt 1989). Our current research, therefore, investigates the degree to which various broad neighbourhoods of language processing are affected by these patterns of collocation and semantic prosody identified by corpus linguists. We start from the processing divisions illustrated in Figure 1 – word recognition and lexical access, semantic processing, and speech production – and we determine whether these are separately sensitive (1) to particular patterns of collocation, and (2) to the abstract generalizations of semantic prosody, in order to determine the psycholinguistic reality of these textual phenomena. The enterprise as a whole is too large to be able to report here. In this first report we therefore restrict ourselves to initial processes of language recognition, particularly visual word recognition and access to the lexicon.

The collocations we investigated stemmed from recent corpus analyses by Kennedy (2003, 2005). Kennedy (2003) analyzed amplifier patterns, the particular ways in which adverbs of degree modify adjectives and verbs, in the British National Corpus (BNC). His research clearly demonstrated that adjectives are very restrictive in their selection of particular boosters and maximizers, as shown in the following examples:

| | | | |
|---|---|---|---|
| ✓ | *absolutely diabolical* | × | *absolutely fledged* |
| ✓ | *fully fledged* | × | *fully blameless* |
| ✓ | *entirely blameless* | × | *entirely diabolical* |
| ✓ | *badly mauled* | × | *badly engrained* |
| ✓ | *deeply engrained* | × | *deeply apposite* |
| ✓ | *particularly apposite* | × | *particularly mauled* |

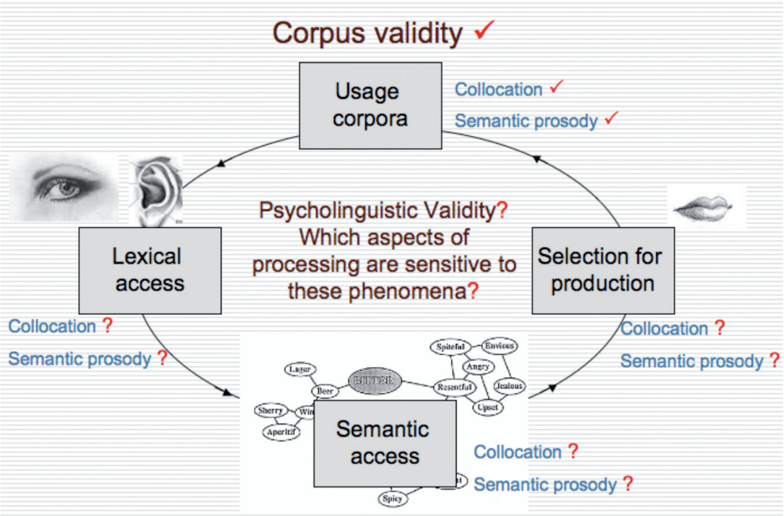


**Figure 1.** The bounds of investigation: To what extent are these different psycholinguistic processes sentitive to the separate corpus-valid phenomena of collocation and semantic prosody?

Kennedy (2005) analyzed the collocations of high frequency English lexical verbs and demonstrated that they too are highly selective in the types of objects they take, thus for example we *end war* but not *finish war*, we *start afresh* but not *begin afresh*, we *lose weight* but don't *receive* it, and *receive support* not *lose* it. We selected some of these linguistic patterns as stimuli and, as described in the method section of Experiment 1, assessed their degree of collocation using standard corpus statistical measures, so to determine whether collocation strength affected fluency of processing in word recognition.

Our study of semantic prosody was grounded in Kjellmer[2] (2005) whose analyses of patterns in the BNC allowed him to identify twenty English verbs that were strongly negative in their semantic prosody (e.g. *cause*: something causes an accident/catastrophe/other negative outcome) and twenty strongly positive verbs (e.g. *achieve*: one achieves objectives/goals/success/other positive outcomes). We took these stimuli and operationalized various measures of direction and strength of semantic prosody, as described in the method section of Experiment 2, so to determine the degree to which fluency of lexical access is affected by prosodic valence.

[2] We are extremely grateful to Göran Kjellmer and Graeme Kennedy for making the results of their corpus analyses available for use in this study, to Carson Maynard for running Experiment 1, and to Gregory Garretson, Graeme Kennedy, the ELI research sharing group, and attendees at ELeGI 2006 for comments on this work.

Our measure of word recognition and access to the mental lexicon was based upon the lexical decision task. This involves participants being shown strings of letters on different trials (for example, *cealt, bread, match, sprong, solp*), and required to indicate whether each letter string is a word or not by pressing the appropriate response key (n, y, y, n, n), with their accuracy and response latency being recorded. A correct "yes" response minimally requires the letter string to be recognized as a word. Meyer and Schvaneveldt (1971) used a variant of this task to demonstrate that, when a reader identifies a word in this way, other words also become active in their mental lexicon. Participants were presented two strings of letters simultaneously, with one string displayed visually above the other. They were required to respond "yes" if both strings were words, otherwise "no". "Yes" responses were about 85 milliseconds faster for pairs of commonly associated words than for pairs of unassociated words, for example, when the word *nurse* appeared above the word *doctor*, participants were faster to respond than when the word *butter* appeared above *doctor*. The fact that response times were facilitated suggested that there was *spreading activation* in memory, where activating the first word's entry results in activation of neighbouring (related) entries, such that the second word is accessed faster to the extent that it is related to the first because it is already partially activated (activation of *bread* spreads to *butter*, but activation of *nurse* does not). Subsequent research has concerned whether these effects are semantic or lexical, i.e. whether automatic priming reflects the retrieval of semantic information, as opposed to the associative/collocational relationships between words (e.g., Williams 1996).

The notion of spreading activation is relatively foreign to most corpus linguists, as is that of semantic prosody to most psycholinguists. Yet these concepts clearly overlap. The question of whether the association underlying spreading activation is lexical or conceptual relates to the question of whether it is syntagmatic or paradigmatic. Equally, effects upon lexical access of collocation *but not* semantic prosody would imply that spreading activation is specific to particular lexical items, whereas effects of collocation *and* semantic prosody would support the notion of generalizations over types. It is possible that lexical recognition mechanisms are sensitive to lexical-level collocation usage alone, and that the generalizations of semantic prosody only show their effect further down the processing stream at semantic access and processing for meaning. Equally, it is possible that there are top-down effects of semantic prosody upon lexical identification. This research assesses these alternatives.

In summary, our specific goals are as follows:

Corpus analyses of language texts demonstrate two phenomena of lexical association: (1) The phenomenon of *collocation*, the co-occurrence of particular words. (2) The phenomenon of *semantic prosody*, whereby a word can be associated with generalized types of words, for example verbs with negative rather than positive objects. In Experiment 1 we determine whether word recognition/lexical access is sensitive to collocation frequency. In Experiment 2 we determine whether it is affected by semantic prosody.

## 2. Experiment 1: The effects of collocation upon lexical access

Experiment 1 is designed to test whether word recognition/lexical access is sensitive to collocation frequency.

### 2.1 Method

*Participants*
This experiment involved 15 adult volunteers recruited from the student population of the University of Michigan, Ann Arbor. They were native speakers of English. They were paid $10 for their participation.

*Materials*
The aim of the experiment was to investigate the extent to which native language users have implicit knowledge of collocation frequency that is brought to bear in word identification and lexical access. Booster and maximizer collocations identified by Kennedy (2003) (e.g. *absolutely diabolical, entirely blameless, badly mauled, deeply engrained*) were kept as pairs or re-sorted as control items which contained the same words combined randomly, thus denying the sequential distribution of English usage (e.g. *absolutely refitted, entirely fledged, badly demarcated, deeply varied*). We then checked the frequency of all of these pairs in the BNC using Mark Davies' VIEW interface (Davies 2007). The complete listing of these items is shown in Appendix 1. Note that the re-sorting occasionally chanced upon a combination which was to be found in the BNC (e.g. *totally disgraceful*) although the collocation frequencies were much higher for the target set. These items constituted the 106 stimulus pairs where both items were words, requiring a "y" response. They were matched with 106 other pairs where either the first (e.g. *veave lessened, screfts engrained*) or second (e.g. *severely swoost, terribly peathed*) item was a non-word. The non-words were selected from the ARC non-word database (Rastle, Harrington and Coltheart 2002) to be between 4 and 8 letters long and to accord with the spelling patterns of English.

Verb object collocations identified by Kennedy (2005) were dealt with in a similar way. We took a set of semantically related verbs for initiation and termination (*start, begin, end, finish, stop*) and selected two high collocates from the BNC (e.g. *end war, end now, stop short, stop wingeing*). The same was done for verbs of transfer (*lose* and *receive*). These natural usages formed the collocations set. We also re-sorted these items to give pairs that, while both words, were not high in collocation strength (e.g. *stop afresh, stop stalemate*). The complete listing of all two-word pairs is given in Appendix 2 along with their BNC frequencies of co-occurrence. As with the boosters and maximizers, these 98 items were matched with 98 pairs where one of the items was a non-word.

*Procedure*

A lexical decision task (Meyer and Schvaneveldt 1971) was used to measure the speed with which participants judged a pair of letter strings as either both words or not. The task was programmed in E-prime (Schneider, Eschman and Zuccolotto 2002) running under Windows XP OS on standard desktop PCs. SuperLab response boxes were used as the input device, allowing participants' reaction times to be recorded with millisecond accuracy.

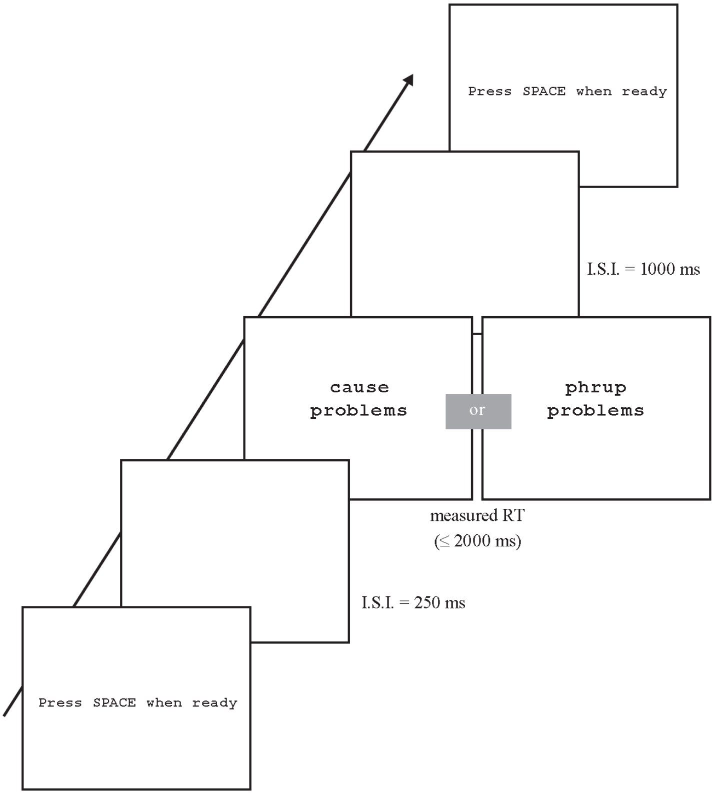


**Figure 2.** Sequence of presentation in lexical decision task

The trial sequence is illustrated in Figure 2. On each trial, the two letter strings appeared in black in the middle of an otherwise white screen, one above the other. Participants were instructed that they would see two strings of letters on the computer screen and they were to judge whether they were both words (yes) or not (no) by pressing either the "y" or "n" button on the response box. They were instructed to respond as quickly and accurately as possible and were given a maximum of 2000 ms to make a decision. After each judgment, a blank screen appeared for 1000 ms followed by a screen reading "Press SPACE key when ready". When participants pressed the space bar to continue, an additional 250 ms gap preceded the next prime-target presentation pair. This procedure allowed participants to take breaks whenever they needed during the flow of the test. Individual reaction times (to the nearest ms following the onset of the letter strings) and accuracy were recorded.

There was an initial practice session of 12 trials in order for the participants to familiarize themselves with the task. After the practice session, the instructions were repeated, and the main session followed with all 408 pairs being presented individually in an individually randomized order of presentation.

*Results*

It is just the results for "yes" trials, where both items were words, that inform the issue of successful lexical access for the pairs. Overall accuracy was good at 91%. We analyzed the response times for correct trials. Outliers (individual responses

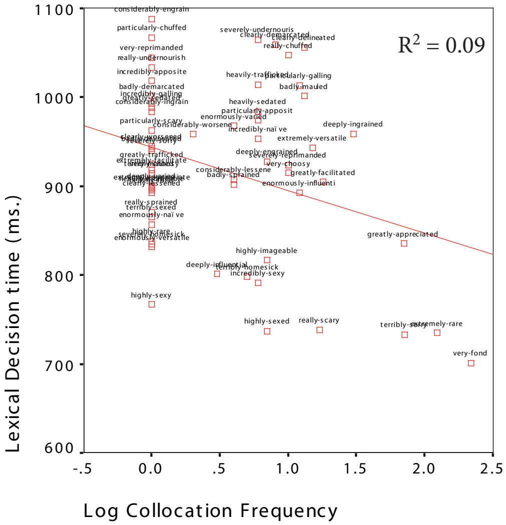


**Figure 3.** Mean judgment time to decide that both letter strings are words as a function of log collocatication frequency of occurrrence of the maximizes and boosters in the BNC

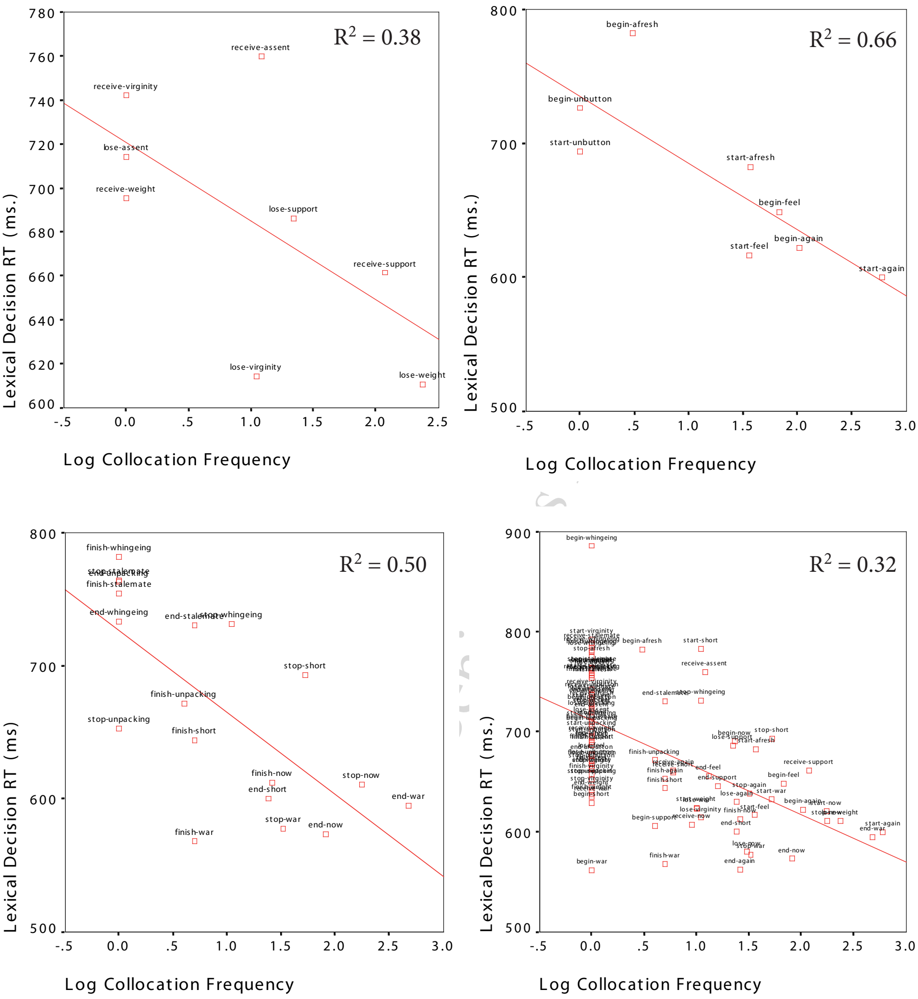


**Figure 4.** Mean judgment time to decide that both letter strings are words as a function of log frequency of occurrence of the verb collocations in the BNC

faster or slower than the participant's mean response time +/– 1.96 standard deviations) were replaced by the participant's mean reaction time. We then calculated the mean response time for each word pair over the 15 participants. These are plotted against the log frequency of the collocation in the BNC for various subsets of the booster and maximizer data in Figure 3 and for the verb data in Figure 4.

It is clear that in every contrast there is a tendency whereby the higher the collocation strength in the language, the faster the participants are able to recognize that both of them are words. Linear regressions predicting response time as a function of log collocation strength explain about 4% of the response time variance for the maximizers and 10% for the boosters. Collocation strength also explains 38% of the variance for the *lose-receive* set, 66% for the *start-begin* set, 50% for the *stop-end-finish* set, and 32% of all verbs combined.

*Conclusion*

Language processing in this lexical decision task is clearly sensitive to patterns of usage of particular collocations. This is so for booster and adjective, maximizer and adjective, and verb and object collocations. Given that the lexical decision task minimally requires word recognition and access to the lexicon, we must conclude that these processes are tuned by experience of particular collocations in usage, so that higher frequency collocations are more readily perceived than lower-frequency ones. The language recognition system has tallied (Ellis 2002a) the co-occurrence of these particular words in prior usage and so tuned itself accordingly to preferentially process them as collocations on future encounters. But what of generalization from these particular patterns to the more schematic associations of semantic prosody? Experiment 2 investigates this.

## 3. Experiment 2: The effects of semantic prosody upon lexical access

Experiment 2 is designed to test whether word recognition/lexical access is sensitive to semantic prosody. As in Experiment 1, a lexical decision test is used to assess whether native speakers are faster to judge that two letter strings are both words if they comprise a semantically prosodic verb paired with an object that matches its valence than if the verb and object are mismatching in prosody. If lexical recognition processes are sensitive to semantic generalizations, then positive valence words (e.g. *goals, maturity, good, benefit*) should be processed faster after positive prosody verbs such as *attain* or *lack* than after negative prosody verbs like *cause* or *provoke*, and, conversely, negative valence words (e.g., *problems, damage, bad, harm*) should be processed faster after negative prosody verbs than after positive prosody verbs.

### 3.1 Method

*Participants*

This experiment involved 15 adult volunteers recruited from the student population of the University of Michigan, Ann Arbor. They were native speakers of English aged between 20 and 30 years. They were paid $10 for their participation.

*Materials*

Verbs judged to have strong positive and negative semantic prosody were selected for the study. Kjellmer (2005) investigated the patterns of 20 positive and 20 negative semantically prosodic verbs and described methods of determining their degree by considering the most frequent nouns associated with them. After he kindly sent us a draft list of these, we developed further operationalizations as follows. Each usage of these verbs was determined in the British National Corpus (BNC) using Mark Davies' VIEW interface (http://view.byu.edu/). The steps were as follows: (1) All collocates were extracted using a 3 slot window to the right and the VIEW pattern *target-verb*.[v*] + noun.all slot 0–3. We recorded the frequencies of the verb, the frequencies of the words with which it collocated, and the frequencies of the particular collocations themselves. We ordered the latter by decreasing frequency. (2) For all collocations with token frequency ≥ 2, or the top 500 most frequent of these if more than that, two independent raters judged each collocate for whether they thought it was positive (P), neutral (.) or negative (N). These raters, the second and third authors of this study, were undergraduates studying psychology, linguistics, and anthropology. Interpretation of words out of context is variable, this indeed is the central theme of the Idiom Principle and of constructional approaches to language, thus there was some variability in these judgments. Nevertheless, the two raters showed enough accord to warrant continuation: the inter-rater agreement was 79% for the positive items, and 85% for the negative items. For each verb we then summed the number of positive, negative, and neutral collocates and computed a variety of indices of prosodic valence and strength, including nP types (the type frequency of the verb's positive associations), %P types (the percentage of collocate types which were positive [nP/(nP+n.+nN)], RatioP/N types (the ratio of nP/nN). Pooling these various indices, we selected ten strongly positively semantically prosodic verbs of the original verb set: *restore, attain, live, achieve, guarantee, advise, grant, gain, regain, lend*, and ten strongly negative verbs: *wreak, inflict, contract, battle, commit, provoke, wage, suffer, cause, cure.* These and their collocation strengths are shown in Table 1.

Each of these 20 verbs were then each combined with various other words to make the "yes" response items in a lexical decision task as in Experiment 1. As shown in Table 2, the paired target items for this task included the two most

**Table 1.** Determination of semantic prosody

| | Frequency (per million words) | | | | | |
|---|---|---|---|---|---|---|
| **Prime** | **Verb** | **All PoS** | **N Collocates (+ or –)** | **% Collocates (+ or –)** | **Ratio + /– Collocates** | **Semantic. Prosody. Valence** |
| attain | 452 | 452 | 41 + | 37 | 13.7 | + |
| cause | 5738 | 12876 | 568 – | 57 | 0.1 | – |
| lack | 1009 | 9871 | 121 + | 41 | 11 | + |
| cure | 521 | 1472 | 55 – | 72 | 0 | – |
| gain | 3663 | 5137 | 316 + | 32 | 5.1 | + |
| suffer | 3421 | 3421 | 400 – | 58 | 0.1 | – |
| guarantee | 1435 | 3911 | 108 + | 30 | 8.3 | + |
| fight | 3871 | 6706 | 194 – | 30 | 0.4 | – |
| grant | 1294 | 7594 | 106 + | 32 | 3.3 | + |
| provoke | 588 | 588 | 74 – | 51 | 0.1 | – |
| restore | 1648 | 1648 | 197 + | 26 | 7 | + |
| encounter | 667 | 1670 | 12 – | 29 | 0.2 | – |
| lend | 1254 | 1254 | 42 + | 24 | 6 | + |
| ease | 1078 | 3020 | 120 – | 49 | 0.1 | – |
| achieve | 6715 | 6715 | 321 + | 32 | 6.2 | + |
| contract | 505 | 11882 | 26 – | 30 | 0.3 | – |
| secure | 2773 | 4548 | 250 + | 32 | 6.4 | + |
| commit | 1339 | 1341 | 78 – | 44 | 0.1 | – |
| emphasize | 654 | 654 | 57 + | 24 | 4.1 | + |
| arouse | 310 | 310 | 26 – | 41 | 0.3 | – |

common collocates of the polarity of the particular prime (e.g., *attain-goals, attain-maturity, cause-problems, cause-damage*), and the two most common collocates of the prime of opposite polarity (e.g., *attain-problems, attain-damage, cause-goals, cause-maturity*). To assess semantic prosody/association rather than specific collocation, each verb was also paired with two generalization items of positive valence (*good* and *benefit*, generating, e.g., the polarity matching *attain-good, attain-benefit* and mismatching *cause-good, cause-benefit*) and two generalization items of negative valence (*bad* and *harm*, generating, e.g., the polarity mismatching *attain-bad, attain-harm*, and matching *cause-bad, cause harm*). This process generated 160 "yes" trials in all.

There was a matching set of 160 "no" trials, 40 where a 4–8 letter non-word from the ARC non-word database was the first item and was paired with the forty collocates, 40 where a non-word preceded the four generalization items (each x 10), and 80 where the 20 verbs were followed by a non-word (each x 4). The

**Table 2.** Prime-target pairings with the top collocates

| | Matched collocates | | Mis-matched collocates | |
|---|---|---|---|---|
| **Prime** | **Target 1** | **Target 2** | **Target 1** | **Target 2** |
| attain | goals | maturity | problems | damage |
| cause | problems | damage | goals | maturity |
| lack | confidence | resources | problems | disease |
| cure | problems | disease | confidence | resources |
| gain | access | understanding | loss | damage |
| suffer | loss | damage | access | understanding |
| guarantee | success | safety | war | battle |
| fight | war | battle | success | safety |
| grant | permission | relief | crisis | violence |
| provoke | crisis | violence | permission | relief |
| restore | confidence | pride | problems | difficulties |
| encounter | problems | difficulties | confidence | pride |
| lend | hand | support | pain | burden |
| ease | pain | burden | hand | support |
| achieve | success | growth | cancer | disease |
| contract | cancer | disease | success | growth |
| secure | knowledge | access | suicide | offence |
| commit | suicide | offence | knowledge | access |
| emphasize | importance | value | suspicion | controversy |
| arouse | suspicion | controversy | importance | value |

experiment as a whole thus comprised 320 trials presented in an individualized random order for each participant to avoid potential order effects.

### *Procedure*

The same lexical decision paradigm as in Experiment 1 was used to measure the speed with which participants judged these pair of letter strings as either both words or not. As in Experiment 1, this allowed us to see if participants judged collocations that they had experienced before faster than novel pairings. In addition, the inclusion of the generalization items allowed the assessment of whether they responded faster when a target word was matched with a verb of the appropriate valence of semantic prosody than with a mismatching one. If word recognition and lexical access is sensitive to the generalizations of semantic prosody/association, then positive valence words (e.g. *goals, maturity, good, benefit*) should be processed faster after positive prosody verbs such as *attain* or *lack* than after negative prosody verbs like *cause* or *provoke*, and, conversely, negative valence words (e.g., *problems, damage, bad, harm*) would be processed faster after negative prosody verbs than after positive prosody verbs.

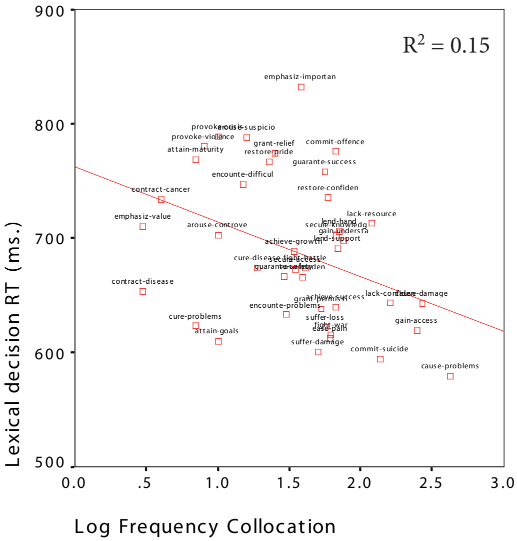


**Figure 5.** Mean judgment times to decide that both letter strings are words as a function of log collocation frequency of occurrence of the *particular* verb-argument collocations in the BNC as assessed in Experiment 2

## 3.2 Results 1. Specific collocations

Consider first the particular collocations shown in the left hand side of Table 2. The participant's reaction times (RT) for correct decisions were analyzed as in Experiment 1. Figure 5 shows the relation between the mean reaction time for deciding that both letter strings of a particular collocation are words and the log frequency of the collocation in the BNC. There is a negative correlation such that the higher the frequency of collocation usage, the faster the judgment. The linear regression predicting RT as a function of log collocation frequency explains 15% of the overall variance.

Thus, as in all of the analyses of Experiment 1, there is clear evidence that the language recognition system is tuned to preferentially process the particular collocations used in this study. But what of generalization?

## 3.3 Results 2. Semantic generalizations

Each of the verbs listed in Table 1 was tested in trials where they were paired with each of the generalization items *good*, *benefit*, *bad* and *harm*, the prediction being that if the language recognition system is sensitive to semantic prosody then

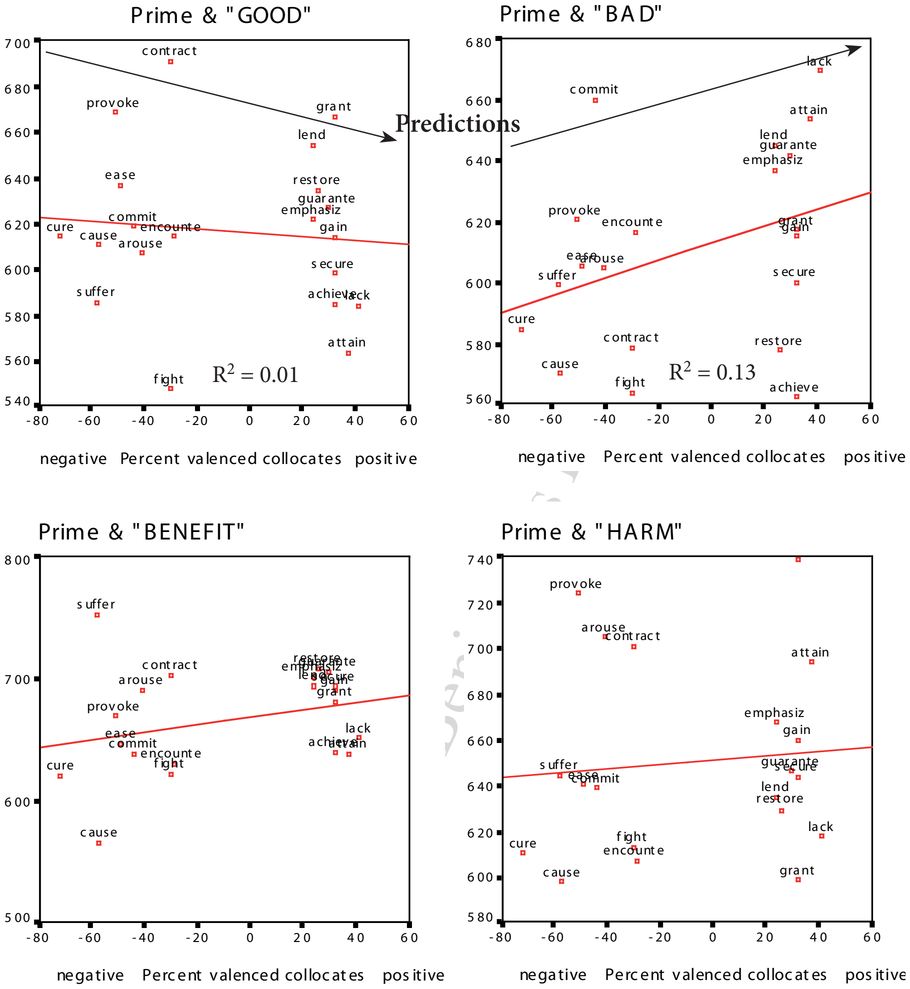
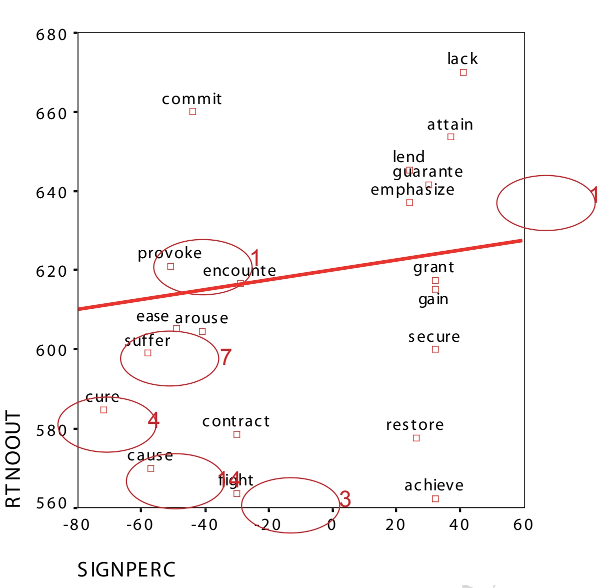


**Figure 6.** Mean judgment times to decide that both letter strings (the prime verb and the target generalization item illustrated) are words as a function of strength (%) of negative or positive prosody as assessed in Experiment 2

the more positive verbs are in their semantic prosody, the faster they should be judged as words when paired with *good* and *benefit*, and the slower when paired with *bad* and *harm*. Figure 6 shows these predicted patterns above the observed mean judgment RTs for each of these pairs as a function of the verb's strength of positive or negative prosody as operationalized using the percentage measure.



Controlling for the actual collocations, $R^2 = 0.01$

**Actual collocations**

**Figure 7.** Mean judgment times for the verbs collocating with "BAD" from the top right panel of Figure 6 showing the actual collocations and their frequencies in the BNC. When these are controlled for, the actual amount of variance additionally explained by strength of prosody is only 1%

Do the results confirm the predicted pattern? The correlation between strength of semantic prosody and judgment time is in the right direction with *good* but explains only 1% of the variance; it is in the right direction with *bad* and explains 13% of the variance; it is in the wrong direction with *benefit* explaining 9% of the variance; it is in the right direction with *harm* but again only explains 1% of the variance. The effects with *good* and *harm* are clearly lacking in substance. The 9% effect with *benefit* is in the wrong direction. The only hope for the hypothesis lies with the effects of semantic prosody when paired with *bad*. But further consideration reveals that to be spurious. The word *bad* was chosen as the clearest negative-polarity word that could be used to test generalization. But scrutiny in the BNC of the verbs we used shows that some of them do form particular collocations with *bad*. These are illustrated in Figure 7. Using the same search window of three words to the right, we find that the English language does evince literal cause, suffer, cure, and fight of *bad* things on occasion, and it looks like these higher frequency collocates are torquing down the regression line towards them. A regression analysis predicting RT judgment of prime + *bad* as a function of log collocation frequency explains a substantial 19% of the variance in its own right. When this is controlled by entering it first into a multiple stepwise regression and then seeing if strength of semantic prosody explains any additional variance, the 13% of variance explained in Figure 6 reduces to an insubstantial 1% in Figure 7.

In sum, using a lexical decision paradigm, we are left with no evidence of an effect of semantic prosody upon word recognition fluency and lexical access.

## 4. Conclusions

These results of Experiments 1 and 2 showed the language recognition system to be tuned to preferentially process frequent verb-argument and booster/maximizer-adjective collocations. Native speakers process familiar collocations more fluently. Just as word recognition is sensitive to the frequency of particular words, particular bigrams, trigrams and other orthographic patterns, particular regularities of spelling-sound correspondence, and other particular patterns in the input (see Ellis 2002a for review), so it is sensitive to word sequences that have become common in the user's usage experience. But the same lexical decision paradigm that so readily shows sensitivity to these patterns of actual collocation usage fails to demonstrate generalization. There is nothing in Experiment 2 to demonstrate that positive valence words (e.g. *goals, maturity, good, benefit*) are processed faster after positive prosody verbs such as *attain* or *lack* than after negative prosody verbs like *cause* or *provoke*, or, conversely, that negative valence words (e.g., *problems, damage, bad, harm*) are processed faster after negative prosody verbs than after positive prosody verbs, unless these word pairs have been previously experienced in *particular* collocations.

It appears then that fluent lexical access is due to memory for particular lexical associations – there are no top-down semantic generalizations upon this level of processing. Meyer and Schvaneveldt (1971) coined the term "semantic priming" to describe the finding of spreading activation in their lexical decision task for associated words such as *doctor-nurse*. It was a plausible appellation at the time. But just as doctors and nurses work together in real life, so they do in the language that describes it, and thus they occur together as collocations in texts. In the light of the demonstration here of robust fluency for particular associates, but not for semantic generalizations, we believe it more appropriate to view these effects upon lexical access as yet other examples of "repetition priming".

Nevertheless, that fluent native speakers show no effects of semantic prosody or semantic association in the recognition processes involved in lexical access does not entail that these phenomena have no effect in other aspects of processing. Indeed, because language texts derive from language users, any distributional systematicities in text must at least entail distributional sensitivity in language users' production processes. Our other investigations (Ellis, Frey and Jalkanen

2007) of post-lexical processing using affective priming paradigms do indeed confirm psycholinguistic effects of semantic prosody, but they suggest that the earliest cognitive locus where these are to be found is in semantic access.

*Language comprehension*

We are by no means the only psycholinguists showing that language comprehension is sensitive to collocation. McDonald and Shillcock (2004) used eye movement recording to reveal that the reading times of individual words are affected by the transitional probabilities of the lexical components. So with sentences like *One way to avoid confusion/discovery is to make the changes during the vacation*, readers read high transitional probability sequences such as *avoid confusion* faster than low transitional probability like *avoid discovery*. In a tightly controlled study, Reali and Christiansen (2007) used both offline and online measures to show that the processing of pronominal object relative clauses was affected by the frequency of co-occurrence of the collocation chunks which formed the clause, so, higher frequency word chunks (The detective who the attorney who *I met* distrusted sent a letter on Monday night) are processed for meaning faster than lower frequency sequences (The detective who the attorney who *I distrusted* sent a letter on Monday night). There is extensive evidence of language users' sensitivity to formulaic sequences in a wide variety of comprehension tasks (Ellis, Simpson-Vlach and Maynard in preparation; Schmitt 2004; Simpson and Ellis 2005). Such results support constructivist views of language whereby frequency of co-occurrence influences the chunking mechanism (Ellis 1996, 2003; Newell 1990) by which multiword units become fused into processing units that are easier to access.

*Language production*

Output production processes are sensitive to collocation too. Schooler (1993) collected likelihood ratio measures of association between various words in order to assess the effect of collocation on memory and processing for recognition and production, showing that word fragment completion was faster for the second word of a strong context collocation (as in *profound-ign*____?) than when the word was shown alone (*ign*____?). Indeed this sensitivity can be shown to be extremely extensive and fine-tuned. Jurafsky, Bell, Gregory and Raymond (2001) analyzed the pronunciation time of successive two-word sequences in the Switchboard corpus to show that in production, humans shorten words that have a higher contextualized probability. The phenomenon is entirely graded with the degree of reduction a continuous function of the frequency of the target word and the conditional probability of the target given the previous word. They argue on the basis of this evidence that the human production grammar must store probabilistic relations between words. As Bybee (2005) quips (after Hebb's (1949) "Cells that fire together, wire together") "Words used together fuse together".

*Language change*
These effects of usage on form play out in language change. Individual learner grammars incorporate variation; this variation changes through use in ways that can lead to the propagation of a change in the speech community that will be established as such in the mental representations of speakers' (variable) grammars, thus resulting in diachronic language change. Bybee (2000; 2002; 2003; 2005; Bybee and Hopper 2001) has developed a model of grammaticization as the process of automatization of frequently-occurring sequences of linguistic elements: (1) Frequency of use leads to weakening of semantic force by habituation; (2) Phonological changes of reduction and fusion of grammaticizing constructions are conditioned by their high frequency; (3) Increased frequency conditions a greater autonomy for a construction, which means that the individual components of the construction (such as *go*, *to* or *-ing* in the example of *be going to*) weaken or lose their association with other instances of the same item (as the phrase reduces to *gonna*); (4) The loss of semantic transparency accompanying the rift between the components of the grammaticizing construction and their lexical congeners allows the use of the phrase in new contexts with new pragmatic associations, leading to semantic change; (5) Autonomy of a frequent phrase makes it more entrenched in the language and often conditions the preservation of otherwise obsolete morphosyntactic characteristics.

*Implications for language: Its usage, processing, learning, and structure*
We process collocates faster and we are more inclined therefore to identify them as units. Such psycholinguistic validation of Firth's maxim (see above) has profound consequences for our understanding of language as a dynamic system (Bybee and Hopper 2001; de Bot, Lowie and Verspoor 2007; Ellis 2007, 2008; Ellis and Larsen Freeman 2006; Larsen-Freeman 1997; MacWhinney 1999) wherein we cannot separate language use from language processing from language learning from language structure from language change:

- One implication for our understanding of language users is that they have an extensive implicit knowledge of particular language sequences (Ellis 2002a).
- One implication for our understanding of psycholinguistics is that both the mental lexicon (Elman 2004) and the mental grammar (Spivey 2006) must be viewed as entirely dynamic and contextualized, with processing being sensitive to these sequential dependencies (Ellis, 2008, Christiansen and Chater 2001; Seidenberg and MacDonald 1999).

- One implication for our understanding of learning is that usage shapes our mental construction of language (Goldberg 2006; Hoey 2005; Langacker 2000; Robinson and Ellis 2007; Tomasello 2003).
- Firth's major legacy, the Forth of Firth, concerns our understanding of language itself. His observations have seeded, over the the last fifty years, a variety of schools of corpus, cognitive, functional, and constructivist linguistics. At their common core is the realization that lexis and grammar are inseparable.

## References

Altman, G. T. 1997. *The Ascent of Babel*. Oxford: OUP.

Becker, A. L. 1983. Toward a post-structuralist view of language learning: A short essay. *Language Learning* 33: 217–220.

Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. 1999. *Longman Grammar of Spoken and Written English.* Harlow: Pearson Education.

Brazil, D. 1995. *A Grammar of Speech*. Oxford: OUP.

Bybee, J. 2000. Mechanisms of change in grammaticalization: The role of frequency. Ms.

Bybee, J. 2002. Phonological evidence for exemplar storage of multiword sequences. *Studies in Second Language Acquisition* 24(2): 215–221.

Bybee, J. 2003. Sequentiality as the basis of constituent structure. In *The Evolution of Language out of Pre-Language*, T. Givón & B. F. Malle (eds), 109–132. Amsterdam: John Benjamins.

Bybee, J. 2005. From Usage to Grammar: The Mind's Response to Repetition. Paper presented at the Linguistic Society of America, Oakland CA.

Bybee, J. & Hopper, P. (eds). 2001. *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins.

Christiansen, M. H. & Chater, N. (eds). 2001. *Connectionist Psycholinguistics.* Westport CO: Ablex.

Davies, M. 2007. *View: Variation in English Words and Phrases*.

De Bot, K., Lowie, W. & Verspoor, M. 2007. A dynamic systems theory to second language acquisition. *Bilingualism: Language and Cognition* 10: 7–21.

Ellis, N. C. 1996. Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition* 18(1): 91–126.

Ellis, N. C. 2002a. Frequency effects in language processing: A review with implications for theories ofimplicit and explicit language acquisition. *Studies in Second Language Acquisition* 24(2): 143–188.

Ellis, N. C. 2002b. Reflections on frequency effects in language processing. *Studies in Second Language Acquisition* 24(2): 297–339.

Ellis, N. C. 2003. Constructions, chunking, and connectionism: The emergence of second language structure. In *Handbook of Second Language Acquisition*, C. Doughty & M. H. Long (eds). Oxford: Blackwell.

Ellis, N. C. 2006. Language acquisition as rational contingency learning. *Applied Linguistics* 27(1): 1–24.

Ellis, N. C. 2007. Dynamic systems and SLA: The wood and the trees. *Bilingualism: Language and Cognition* 10: 23–25.

Ellis, N. C. 2008. The dynamics of language use, language change, and first and second language acquisition. *Modern Language Journal* 92(2): 232–249.

Ellis, N. C., Frey, E. & Jalkanen, I. 2007. The psycholinguistic reality of collocation and semantic prosody – neighbourhoods of knowing (2): Semantic access. Paper presented at the UWM Linguistics Symposium on Formulaic Language, University of Wisconsin, Milwaukee, 18–21 April 2007.

Ellis, N. C. & Larsen Freeman, D. 2006. Language emergence: implications for applied linguistics. *Applied Linguistics* 27 (4).

Ellis, N. C., Simpson-Vlach, R. & Maynard, C. In preparation. The processing of formulas in native and second language speakers: Psycholinguistic and corpus determinants. *TESOL Quarterly*.

Elman, J. L. 2004. An alternative view of the mental lexicon. *Trends in Cognitive Science* 8: 301–306.

Erman, B. & Warren, B. 2000. The idiom principle and the open choice principle. *Text* 20: 29–62.

Firth, J. R. 1957a. *A Synopsis of Linguistic Theory: 1930–1955*. Oxford: Basil Blackwell.

Firth, J. R. 1957b. *Papers in Linguistics: 1934–1951*. London: OUP.

Gernsbacher, M. A. 1994. *A Handbook of Psycholinguistics*. San Diego CA: Academic Press.

Goldberg, A. E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: OUP.

Granger, S. & Meunier, F. (eds). 2008. *Phraseology: An Interdisciplinary Perspective*. Amsterdam: John Benjamins.

Harley, T. A. 1995. *The Psychology of Language: From Data to Theory*. Hove: Taylor and Francis.

Hebb, D. O. 1949. *The Organization of Behaviour*. New York NY: John Wiley and Sons.

Hoey, M. 2005. *Lexical Priming: A New Theory of Words and Language*. London: Routledge.

Hunston, S. & Francis, G. 2000. *Pattern Grammar: A Corpus-Driven Approach to the Lexical Grammar of English*. Amsterdam: John Benjamins.

Jurafsky, D., Bell, A., Gregory, M. & Raymond, W. D. 2001. Probabilistic relations between words: Evidence from reduction in lexical production. In *Frequency and the Emergence of Linguistic Structure*, J. Bybee & P. Hopper (eds), 229–254. Amsterdam: John Benjamins.

Kennedy, G. 2003. Amplifier collocations in the British National Corpus: Implications for English language teaching. *TESOL Quarterly* 37: 477–486.

Kennedy, G. 2005. Collocational patterning with high frequency verbs in the British National Corpus. Paper presented at the American Association of Applied Corpus Linguistics conference, University of Michigan, Ann Arbor, 12–15 May 2005.

Kjellmer, G. 1987. Aspects of English collocations. In *Corpus Linguistics and Beyond*, W. Meijs (ed.). Amsterdam: Rodopi.

Kjellmer, G. 2005. Collocations and Semantic Prosody. Paper presented at the American Association of Applied Corpus Linguistics conference, University of Michigan, Ann Arbor, 12–15 May, 2005.

Langacker, R. W. 2000. A dynamic usage-based model. In *Usage-Based Models of Language*, M. Barlow & S. Kemmer (eds), 1–63. Stanford CA: CSLI.

Larsen-Freeman, D. 1997. Chaos/complexity science and second language acquisition. *Applied Linguistics* 18: 141–165.

Leech, L. 2000. Grammars of spoken English: New outcomes of corpus-oriented research. *Language Learning* 50: 675–724.

Levelt, W. J. M. 1989. *Speaking: From Intention to Articulation*. Cambridge MA: The MIT Press.

Louw, B. 1993. Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In *Text and Technology: In Honour of John Sinclair*, M. Baker, G. Francis & E. Tognini-Bonelli (eds), 157–176. Amsterdam: John Benjamins.

MacWhinney, B. (ed.). 1999. *The Emergence of Language*. Hillsdale NJ: Lawrence Erlbaum.

McDonald, S. A. & Shillcock, R. C. 2004. Eye-movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science* 14: 648–652.

Meyer, D. E. & Schvaneveldt, R. W. 1971. Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology* 90: 227–234.

Nattinger, J. R. 1980. A lexical phrase grammar for ESL. *TESOL Quarterly* 14: 337–344.

Newell, A. 1990. *Unified Theories of Cognition*. Cambridge MA: Harvard University Press.

Pawley, A. & Syder, F. H. 1983. Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In *Language and Communication*, J. C. Richards & R. W. Schmidt (eds), 191–225. London: Longman.

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

Reali, F. & Christiansen, M. H. 2007. Word chunk frequencies affect the processing of pronomial object-relative clauses. *Quarterly Journal of Experimental Psychology* 60: 161–170.

Robinson, P. & Ellis, N. C. (eds). 2007. *A Handbook of Cognitive Linguistics and SLA*. Mahwah NJ: Lawrence Erlbaum.

Schmitt, N. (ed.). 2004. *Formulaic Sequences*. Amsterdam: John Benjamins.

Schneider, W., Eschman, A. & Zuccolotto, A. 2002. *E-Prime User's Guide*. Pittsburgh PA: Psychology Software Tools.

Schooler, L. J. 1993. *Memory and the Statistical Structure of the Environment*. Pittsburgh PA: Carnegie Mellon University.

Seidenberg, M. S. & MacDonald, M. C. 1999. A probabilistic constraints approach to language acquisition and processing. *Cognitive Science* 23: 569–588.

Simpson, R. & Ellis, N. C. 2005. An academic formulas list (Afl): Extraction, validation, prioritization. Paper presented at Phraseology 2005, Louvain-la-Neuve, Belgium, 13–15 October 2005.

Sinclair, J. 1991. *Corpus Concordance Collocation*. Oxford: OUP.

Sinclair, J. 2004. *Trust the Text: Language, Corpus and Discourse*. London: Routledge.

Spivey, M. 2006. *The Continuity of Mind*. Oxford: OUP.

Tomasello, M. 2003. *Constructing a Language*. Boston MA: Harvard University Press.

Williams, J. N. 1996. Is automatic priming semantic? *European Journal of Cognitive Psychology* 22: 139–151.

## Appendices

**Appendix 1.** The booster and maximizer collocations tested in Experiment 1

| Collocation | BNC frequency | Re-sort control | BNC frequency |
|---|---|---|---|
| MAXIMIZERS | | | |
| absolutely-diabolical | 9 | absolutely-refitted | 0 |
| completely-refitted | 5 | completely-chuffed | 0 |
| dead-chuffed | 7 | dead-blameless | 0 |
| entirely-blameless | 8 | entirely-fledged | 0 |
| fully-fledged | 50 | fully-diabolical | 0 |
| perfectly-contestable | 17 | perfectly-unsuited | 0 |
| totally-unsuited | 10 | totally-desolate | 0 |
| utterly-desolate | 6 | utterly-contestable | 0 |
| absolutely-knackered | 6 | absolutely-inelastic | 0 |
| completely-inelastic | 9 | completely-proportioned | 0 |
| dead-boring | 14 | dead-fortuitous | 0 |
| entirely-fortuitous | 6 | entirely-knackered | 0 |
| fully-conversant | 23 | fully-boring | 0 |
| perfectly-proportioned | 12 | perfectly-unprepared | 0 |
| totally-unprepared | 17 | totally-disgraceful | 2 |
| utterly-disgraceful | 7 | utterly-conversant | 0 |
| absolutely-gorgeous | 26 | absolutely-outclassed | 0 |
| completely-outclassed | 4 | completely-gorgeous | 0 |
| dead-drunk | 12 | dead-coincidental | 0 |
| entirely-coincidental | 5 | entirely-irresponsible | 2 |
| fully-clothed | 42 | fully-drunk | 1 |
| perfectly-manicured | 6 | perfectly-illegible | 0 |
| totally-illegible | 5 | totally-clothed | 0 |
| utterly-irresponsible | 4 | utterly-manicured | 0 |
| BOOSTERS | | | |
| badly-mauled | 12 | badly-demarcated | 0 |
| clearly-demarcated | 7 | clearly-lessened | 0 |
| considerably-lessened | 3 | considerably-engrained | 0 |
| deeply-engrained | 6 | deeply-varied | 0 |
| enormously-varied | 3 | enormously-versatile | 0 |
| extremely-versatile | 14 | extremely-facilitated | 0 |
| greatly-facilitated | 17 | greatly-trafficked | 0 |
| heavily-trafficked | 5 | heavily-imageable | 0 |
| highly-imageable | 6 | highly-sexy | 0 |
| incredibly-sexy | 5 | incredibly-galling | 0 |
| particularly-galling | 11 | particularly-chuffed | 0 |

| | | | |
|---|---|---|---|
| really-chuffed | 9 | really-undernourished | 0 |
| severely-undernourished | 5 | severely-homesick | 0 |
| terribly-homesick | 4 | terribly-choosy | 0 |
| very-choosy | 9 | very-mauled | 0 |
| badly-sprained | 3 | badly-delineated | 0 |
| clearly-delineated | 12 | clearly-worsened | 0 |
| considerably-worsened | 1 | considerably-ingrained | 0 |
| deeply-ingrained | 29 | deeply-influential | 2 |
| enormously-influential | 11 | enormously-naïve | 0 |
| extremely  rare | 122 | extremely-appreciated | 0 |
| greatly-appreciated | 69 | greatly-sedated | 0 |
| heavily-sedated | 5 | heavily-fond | 0 |
| highly-sexed | 6 | highly-rare | 0 |
| incredibly-naïve | 5 | incredibly-apposite | 0 |
| particularly-apposite | 5 | particularly-scary | 0 |
| really-scary | 16 | really-sprained | 0 |
| severely-reprimanded | 9 | severely-sorry | 0 |
| terribly-sorry | 70 | terribly-sexed | 0 |
| very-fond | 216 | very-reprimanded | 0 |

**Appendix 2.** The verb object collocations tested in Experiment 1

| Collocation | Condition | BNC freq | Collocation | Condition | BNC freq |
|---|---|---|---|---|---|
| start-again | colloc | 599 | begin-virginity | control | 0 |
| start-afresh | colloc | 36 | begin-support | control | 3 |
| start-feel | control | 35 | begin-assent | control | 0 |
| start-unbutton | control | 0 | end-again | control | 25 |
| start-war | control | 51 | end-afresh | control | 0 |
| start-stalemate | control | 0 | end-feel | control | 12 |
| start-now | control | 171 | end-unbutton | control | 0 |
| start-unpacking | control | 0 | end-war | colloc | 478 |
| start-short | control | 10 | end-stalemate | colloc | 4 |
| start-wingeing | control | 0 | end-now | control | 81 |
| start-weight | control | 9 | end-unpacking | control | 0 |
| start-virginity | control | 0 | end-short | control | 23 |
| start-support | control | 0 | end-wingeing | control | 0 |
| start-assent | control | 0 | end-weight | control | 0 |
| begin-again | control | 102 | end-virginity | control | 0 |
| begin-afresh | control | 2 | end-support | control | 15 |
| begin-feel | colloc | 67 | end-assent | control | 0 |
| begin-unbutton | colloc | 0 | finish-again | control | 4 |
| begin-war | control | 0 | finish-afresh | control | 0 |

| | | |
|---|---|---|
| begin-stalemate | control | 0 |
| begin-now | control | 22 |
| begin-unpacking | control | 0 |
| begin-short | control | 0 |
| begin-wingeing | control | 0 |
| begin-weight | control | 0 |
| finish-short | control | 4 |
| finish-wingeing | control | 0 |
| finish-weight | control | 0 |
| finish-virginity | control | 0 |
| finish-support | control | 0 |
| finish-assent | control | 0 |
| stop-again | control | 31 |
| stop-afresh | control | 0 |
| stop-feel | control | 0 |
| stop-unbutton | control | 0 |
| stop-war | control | 32 |
| stop-stalemate | control | 0 |
| stop-now | control | 174 |
| stop-unpacking | control | 0 |
| stop-short | colloc | 52 |
| stop-wingeing | colloc | 10 |
| stop-weight | control | 0 |
| stop-virginity | control | 0 |
| stop-support | control | 0 |
| stop-assent | control | 0 |
| lose-again | control | 23 |
| lose-afresh | control | 0 |
| lose-feel | control | 0 |
| lose-unbutton | control | 0 |
| lose-war | control | 9 |
| finish-feel | control | 0 |
| finish-unbutton | control | 0 |
| finish-war | control | 4 |
| finish-stalemate | control | 0 |
| finish-now | colloc | 25 |
| finish-unpacking | colloc | 3 |
| lose-stalemate | control | 0 |
| lose-now | control | 29 |
| lose-unpacking | control | 0 |
| lose-short | control | 0 |
| lose-wingeing | control | 0 |
| lose-weight | colloc | 236 |
| lose-virginity | colloc | 10 |
| lose-support | control | 21 |
| lose-assent | control | 0 |
| receive-again | control | 5 |
| receive-afresh | control | 0 |
| receive-feel | control | 0 |
| receive-unbutton | control | 0 |
| receive-war | control | 0 |
| receive-stalemate | control | 0 |
| receive-now | control | 8 |
| receive-unpacking | control | 0 |
| receive-short | control | 5 |
| receive-wingeing | control | 0 |
| receive-weight | control | 0 |
| receive-virginity | control | 0 |
| receive-support | colloc | 117 |
| receive-assent | colloc | 11 |