An acquired taste: How reading literature affects sensitivity to word distributions when judging literary texts

Justine Kao, Robert Ryan, Melody Dye & Michael Ramscar
Department of Psychology, Stanford University,
Jordan Hall, Stanford, CA 94305

Abstract

This study examines how reading habits affect people's sensitivity to word distributions in literary and non-literary writing. We manipulated eight literary and non-literary passages, creating modified versions that had lower word chunk frequencies but higher individual word frequencies than the originals. Subjects were then asked to rate the passages' quality of writing. Results showed that subjects with more experience reading literary writing (literary readers) gave higher ratings to original literary passages, while subjects with less literary reading experience (non-literary readers) preferred modified versions. Subjects with both types of reading habits rated original versions of non-literary passages higher. This indicates that literary readers are sensitive to frequencies of word chunks containing words that appear more frequently in the literary genre, while non-literary readers are not. We suggest that, over time, people can acquire slightly different representations of the probabilistic structure of language through their specific linguistic exposure.

Keywords: Psycholinguistics, Corpus linguistics, Word distributions, Genre differences, Reading habits, Discourse processes, Literary studies

Introduction

With one careful, calculated sip, a wine connoisseur can detect the subtle differences of quality between wines, and may even note the year and vineyard in which the grapes were grown. We, on the other hand, may stumble upon a thirty-year-old Bordeaux and not be able to tell it apart from a ten-dollar bottle. Appreciation for wine, like appreciation for high fashion or opera, is an acquired skill. Many fine things in life require years of experience to generate true appreciation. In what follows, we ask whether or not this "connoisseur phenomenon" translates to appreciation for literature as well. Is the ability to detect skill and beauty in literature also an acquired taste? If so, what is being acquired through the act of reading? Will avid readers have a stronger appreciation for good-quality writing, or be more sensitive to subtle changes of word choice?

As writing becomes an increasingly important form of communication and a central aspect of our lives, many studies have been conducted on the ways in which we are affected by what we read. Previous research shows that frequent readers are more sensitive to ambiguities in literary texts and are more likely to provide nuanced interpretations of them than infrequent readers (Dixon et al., 1993). Further, students who read recreationally perform better on reading comprehension and vocabulary tests (Anderson et al., 1988; Cipielwski & Stanovich, 1992), suggesting a relationship between reading enjoyment and competence. Still another study shows that frequent readers of literary writing have higher empathy and social measures than readers of non-literary writing (Mar et al., 2006). These results suggest that certain effects on our social, reasoning, and linguistic skills may be closely connected to the kinds of reading we engage in.

While these studies focus on higher-order social and cognitive effects of reading, we are interested here in examining how reading shapes readers' sensitivity to distributions of words. More specifically, we seek to explore whether readers' experience reading literary or non-literary writing shapes their sensitivities to word and chunk frequencies, and further, whether these fine-tuned sensitivities affect their judgments of quality when rating texts from different genres.

The Probabilistic Nature of Natural Languages

A slew of recent studies have shown that language users are sensitive to the distributional patterns of sounds, words, and even larger linguistic structures such as word sequences, or word 'chunks,' in the language they speak (see e.g., Altmann & Steedman, 1988; Bell et al., 2009; Bod et al., 2003; Bybee, 2002, 2006; Bybee & Hopper, 2001; De Long et al., 2005; Hale, 2003; Levy, 2008; Otten & Van Berkum, 2008; Pierrehumbert 2001, 2003; Ramscar et al., in press; Perruchet & Pacteau, 1990; Servan-Schreiber & Anderson, 1990). In many ways, the idea that we pay attention to how words are used is hardly surprising. It seems obvious, for example, that "a daunting task" sounds more "right," or more familiar, than "a daunting job." In fact, although *job* is a much higher frequency word than *task*, "a daunting task" appears 191 times on the Corpus of Contemporary American English (COCA), while "a daunting job" occurs only 6 times. We are sensitive to the different frequencies of the two chunks and prefer the one with the higher frequency. Since there is no real reason why it is less appropriate to describe a job as daunting, the preference for "a daunting task" over "a daunting job" does not seem to be driven by the appropriateness of the phrases' inherent *meanings,* but rather how the words are usually used.

The reason why we can sense these subtle mismatches is because words do not co-occur with each other with equal frequency. Indeed, the distribution of words in languages is

highly systematic (Baayen, 2001), and listeners are clearly sensitive to how words co-occur in sensible, and less sensible ways (see e.g., Wicha, Bates, Moreno, & Kutas, 2003). These kinds of co-occurrence patterns offer a rich and readily available source of information for anyone learning to understand the way that language relates to the world, and there is considerable evidence to support the idea that people are sensitive to this information.

However, it is critical to note that every person's internal model of his or her language is trained on a slightly different corpus. In other words, each person hears and reads different things throughout his or her life, and over time these differences in the input may result in different representations of the language. In written language, for example, genres of writing have been observed to differ on a number of linguistic dimensions. Research on corpus comparison and genre detection makes use of the idea that word distributions – how words are used and which words are used – differ across genres (Biber 1988, 1993; Eisenbeis & Avery, 1972; Karlgren & Cutting, 1994; Lee & Myaeng, 2002, Xiao & McEnery, 2005). Work in literary theory has also suggested that literary texts often use low-frequency words to foreground certain elements of writing (Miall & Kuiken, 1994, Mukarovský, 1964), while non-literary texts tend to use more conventional words to convey meaning clearly. Given that there is marked variation in the distributions of words that people will be exposed to over the course of their lives, it seems likely that people will have different sensitivities to word distributions depending on their "training sets." We examine this possibility through the lens of writing genres.

'Literary' and 'Non-literary' Words

For our purposes here, we class writing into two primary domains: literary and non-literary. Much of what people read can be identified as one of the two, with fiction and poetry belonging to the former category, and newspaper articles and textbooks to the latter. Based on whether a word occurs more frequently in literary writing or non-literary writing, we can refer to it as a 'literary' word or a 'non-literary' word. For example, "abruptly" is a literary word (37 per million in the fiction corpus and 6.7 per million in the newspaper corpus), while "actively" is a non-literary word (2.54 per million in fiction and 9.97 in newspapers) (Corpus of Contemporary American English (COCA)).

As we will illustrate in a later section, literary texts tend to contain more literary words, while non-literary texts tend to contain more non-literary words. Since literary and non-literary words are *defined* by how often they occur overall in literary and non-literary writing, this may not seem entirely surprising. However, it sheds light on the deeper point that the words in a given piece of writing will have different distributions depending on the corpus you examine (e.g., the frequency and usage of "abruptly" will differ sharply between a "non-literary" newspaper corpus and a "literary" fiction corpus).

This has implications for how people may be affected by their reading practices. Given that some people's reading habits may make them more familiar with one "corpus" than another (i.e., they may be more widely read newspapers and journal articles than fiction and poetry), this difference in exposure should translate into a corresponding difference in their probabilistic representation of the distributions of words within their language. In other words, readers within different genres will have learned somewhat different distributional patterns, and these differences should be similar to the ones that we can actually research and quantify by analyzing different corpora.

This leads to testable predictions. For example, we would expect that literary readers would be more sensitive to the probabilistic distributions of literary words than non-literary readers, and we would also expect them to have a better understanding of the environment – or linguistic context – in which such words are likely to occur. Thus, they should show higher sensitivity than non-literary readers to the frequencies of *chunks* of words in literary texts.

Reading Habits and Judgment: Experiment

In order to test the predictions detailed above, we selected four excerpts of choice contemporary fiction writing and four excerpts of non-literary writing. We then systematically manipulated the frequencies of several chunks (short sequences of words) within each passage, creating modified versions of each of the eight passages. Our method of modification is detailed in the section "Manipulation of passages." After creating the 8 modified versions, we had16 testing passages total: 4 literary and 4 non-literary *original* passages, which contain higher overall chunk frequencies but lower overall word frequencies, and 4 literary and 4 non-literary *modified* passages, which contain lover overall chunk frequencies but higher overall word frequencies. We hope to examine whether subjects' evaluations of writing quality differ for the original and modified versions, and further whether literary and non-literary readers' evaluation of literary and non-literary texts also diverge.

We hypothesize that for literary texts, literary readers will give higher ratings to literary passages containing chunks that have higher frequencies, because these chunks will be more familiar in the corpus they have been trained on, and thus more representative of their internal models of language (e.g., they should recognize "adamantine luster" as a frequent literary pairing and prefer it over "adamantine milk," which is not a frequent literary pairing). By contrast, we hypothesize that non-literary readers, who lack the same levels of exposure to 'literary' words and their contexts, will only be sensitive to individual word frequencies, and will prefer more highly frequent words even when they are used in contexts (e.g., "adamantine milk") that would seem anomalous or even jarring to a literary reader. In terms of quality ratings, this suggests that literary readers will prefer the original literary passages with higher chunk frequencies,

whereas non-literary readers will prefer the modified literary passages with higher individual word frequencies.

With regards to the non-literary texts, the picture is less clear. It may be that we should expect the opposite effect: that literary readers will prefer modified passages while non-literary readers will prefer the originals. However, it also seems likely that our literary readers, who read for pleasure, may read more widely than our non-literary readers, and be sensitive to our non-literary manipulations as well.

Participants

Participants were 31 Stanford University undergraduates recruited for credit for an introductory psychology course. All subjects were monolingual English speakers.

Materials

Four excerpts from literary writing and four excerpts from non-literary writing, each ranging from 80 to 130 words in length, were selected as materials. The literary passages were selected from four separate stories in "*The Vintage Book of Contemporary American Short Stories*," a collection of short stories featuring distinctive short fiction in American English published within the last 25 years. Three journalistic, or non-literary, English passages were selected from articles in the New York Times during the past year, and one non-literary passage was chosen from a reading comprehension article in a 2009 GRE prep book. Passages from each genre varied in style and content. We chose materials from these sources because they reflect high quality of writing, offer a variety of styles and themes, and are not famous or widely enough read to be likely to be recognized by our subjects during the survey.

Methods

Assessment of Passages

To explore the degree to which literary texts tend to contain more literary words and non-literary texts tend to contain more non-literary words, we examined the 400 million word COCA corpus (Davies, 2009) recording the frequency of each word in each passage in the fiction corpus, the newspaper corpus, and the corpus as a whole. The average log frequencies of the passages in the three corpora are shown in figure 1. This analysis revealed that within the specific corpora, the literary passages had significantly higher average frequencies in the fiction corpus than in the newspaper corpus ($t(670)=2.3148$; $p < 0.05$) whereas the average frequencies of the non-literary texts in the fiction and newspaper corpora were not significantly different ($t(584)=-1.0288$; $p>0.05$). This suggests that words occurring in literary texts are more frequent in literary than non-literary texts, while words in non-literary texts are more evenly distributed across literary and non-literary texts. This idea is supported by an analysis of the overall corpus, which revealed the literary passages to have higher average frequencies than the non-literary passages ($t(627)=2.2786$; $p<0.05$). Together these findings suggest that literary texts make specialized use of a specific subset of the overall corpus, rather than employ a markedly different vocabulary. Consistent with this idea, a 2 (literary versus non-literary text) x 2 (fiction versus newspaper corpus) ANOVA of the average frequencies of the texts revealed an interaction between text type and corpus type ($F(1, 627)=13.324$, $p<0.001$), and a main effect of text type ($F(1,627)=121.926$, $p<0.001$).

A more fine-grained analysis of the texts further supported the idea that the distribution of vocabulary items is specialized in different kinds of writing. When we compared the pair-wise frequency of each word in each specific corpus, we found (unsurprisingly) that the pair-wise frequencies of the words in the literary passages were significantly higher in the corpus for fiction writing than in the corpus for newspaper writing ($t(335)=11.4987$; $p<0.001$), but also that the reverse was true for each of the words in non-fiction passages ($t(292)=-4.7295$; $p <0.001$). In other words, what appears to set literary and non-literary writing apart is not that they make use of specialized sets of words, but rather that words are used in specialized ways in different kinds of writing, and, at least in this sample, the distribution of vocabulary within literary writing in English appears to be particularly distinctive.

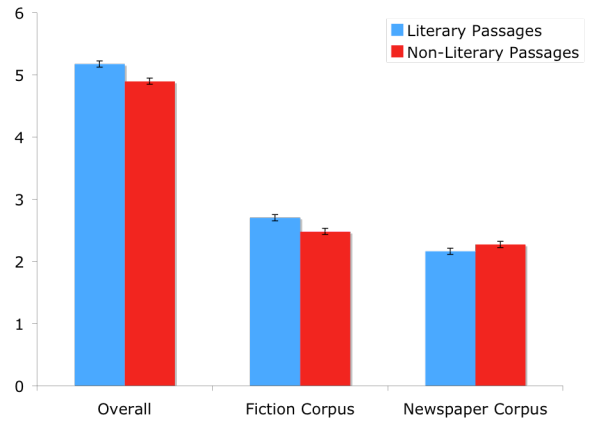

Figure 1. Average log frequencies of passages in different corpora

Manipulation of Passages

After analyzing the passages, we then manipulated the frequencies of three to seven chunks (strings of words) within each literary and non-literary passage, lowering chunk frequency while simultaneously retaining (or even raising) average individual word frequency. Measures of individual word frequency were taken form COCA, while chunk frequency was based on the number of 'hits' a chunk returned on Google. The reason why we used Google to measure chunk frequencies is because its magnitude allows us to find 'hits' for word sequences that are several words long, while many longer word sequences would return 0 counts even in large corpora like COCA, and thus fail to

measure the differences in frequencies of longer word chunks.

In the following example, (a) is the original chunk, and (b) is the modified chunk.

(a) On the further ***side*** of the field[1]
(b) On the further ***part*** of the field

While *side* and *part* have highly similar meanings in this context, *side* has an average frequency of 317/ million in COCA, while *part* has frequency of 479/ million, suggesting that *part* is a more frequent word in English than *side*.

In terms of chunk frequencies, "further side" returns 259,000 hits on Google, whereas "further part" returns 374,000 hits, suggesting that in English, "part" is more likely than "side" given "further" (Miller & Chomsky's[2] famous point about the lack of evidence for specific strings in English can be illustrated by considering that there are insufficient instances of "part of" and "side of" in the 400 million words of COCA to facilitate an analysis). Finally, the highly frequent "of" is the most likely word to follow both "side" and "part" in English, and by both our COCA and Google measures, the likelihood of "of" given "part" is three times that of "side"("part of" has 132176 Google hits compared to 40446 for "side of," and "part of" occurs on average 330 times per million words in COCA, as compared to 101/million for "side of").

Thus, the "average probabilities" of English suggest that, as a string of words, (b) is much more likely than (a). However, the original chunk (a) returned 898 hits on Google, while the modified chunk (b) returned 0 hits. Thus, although and both (a) and (b) appear to be similar in meaning and equally "grammatical," and although the average frequency of all words, and the average transitional probabilities between them in English as a whole are higher in (b), given that its chunk frequency is considerably lower, it appears that the likelihood of actually encountering (b) in English is lower than it is for (a).

Since we wished to manipulate word and chunk frequencies while keeping the meaning of the passages relatively constant, chunks were selected for modification in the manner just described based on whether or not they contained a word that could be replaced with a synonym that had a similar or higher frequency.

Procedure and design

All surveys were designed and distributed using the Qualtrics online survey software. Each survey had four literary and four non-literary passages, half of which were original excerpts, and half of which were modified as described in the section above. Two versions of the survey were distributed: either the odd-numbered passages were modified and the even-numbered passages were kept as the original, or vice versa. Participants were randomly assigned one of the two versions.

[1] Taken from "Emergency," by Denis Johnson
[2] Miller & Chomsky, (1963)

Participants were surveyed individually on a computer. They were asked to read the instructions in the survey carefully, and the time it took each subject to complete the survey was recorded to make sure they spent enough time reading the passages and answering questions.

Participants were presented with each passage in the same order and asked to read carefully. While each participant read the passages in the same order, the order of the passages was counterbalanced with respect to passage type. For example, in the version of the survey in which the odd-numbered passages were modified, the passages appeared in the order of: modified literary, original non-literary, modified literary, original literary, modified non-literary, original non-literary, modified non-literary, original literary. This design should weaken the effects of passage type ordering on subjects' preferences.

After subjects finished reading the passage as a whole, the same passage appeared again, but this time with a selection highlighted. They were asked to rate the quality of the highlighted section on a 7-point scale, with 7 being "Very well-written," and 1 being "Very poorly written." Each passage was equally divided into three sections, separately highlighted and presented to the subjects for rating.

After participants finished reading and rating all eight passages, they were asked to provide an estimate of how many hours a week they usually spent reading literary texts (including poetry, magazine stories, creative non-fiction, and novels) and non-literary texts (including text books, newspaper articles, and academic papers). In order to arrive at a score of how much more experience each subject had reading literary writing compared to reading non-literary writing, the hours reading literary texts was divided by the hours reading non-literary texts, a ration we will refer to as the 'literary reading bias.' We use this as our measure for subjects' reading habits because it reflects the relative amount of time they read literary texts versus non-literary texts, which for our purposes is the salient feature of their reading habits.

Results

A repeated measures ANCOVA of participant ratings of the modified and non-modified passages with literary reading bias as a continuous covariate revealed a significant interaction between literary reading scores and within-genre preference ($F(1, 21) = 3.095$; $p < 0.05$; see figure 2). To facilitate further analysis, subjects were divided equally into two groups based on their literary reading bias, with subjects whose scores were above the median placed in one group, and subjects whose scores were below the median placed in the other. The within-genre preference of each subject was measured using the difference between his or her average ratings for original and modified passages of each genre. The two groups' average preferences within each genre are shown in figure 2.

Further, participants who read more fiction relative to non-fiction writing showed a stronger preference for the

unmodified literary texts compared to participants who read more non-fiction ($t(29) = 1.7377$; $p<0.05$), and a one-sample t-test revealed that participants who read more fiction showed an overall preference for the original literary passages ($t(14)=1.856$; $p<0.05$. For non-literary passages, there was no significant difference between the within-genre preferences of subjects in the two groups ($t(29)=0.6556$; $p>0.5$), and while both groups showed a preference for the original non-literary passages, these preferences were not significant.

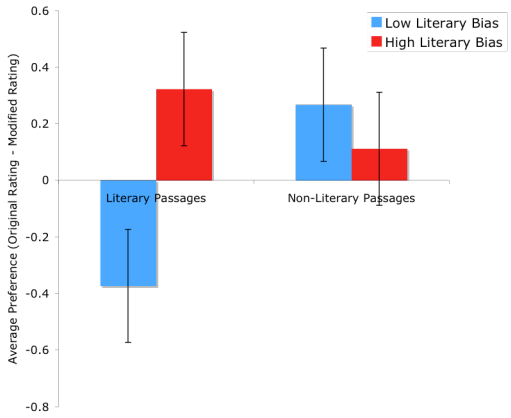

Figure 2. Literary and non-literary readers' preference for original passages in the two genres

Discussion

As predicted, there was a significant interaction between subjects' reading habits and their reading preferences. How might one explain these results?

Only people who are exposed to the distributional properties of those words in literary contexts appear be sensitive to our manipulations, which is consistent with our assessment of the passages, where we found that literary writing uses words in ways that are literary specific. On the other hand, both literary and non-literary readers were sensitive to the manipulation of non-literary passages. One reason may be that non-literary writing makes use of less specialized distributions, as shown in our corpus analysis. Literary writing can be thought of as a specialized form of writing that re-employs and expands upon distributional information also present in non-literary writing, which makes literary readers still reasonably familiar with the distributions of words in non-literary texts, whereas the same cannot necessarily be said for non-literary readers and literary texts. Another reason may be that our social and cultural lives naturally enforce a non-literary expertise on all readers, while literary expertise is more a matter of individual practice.

One potential weakness for our study was that we relied on self-report to measure our subjects' reading habits. There may be issues of accuracy in recall, given that subjects were trying to judge the exact number of hours they spent reading in a given week. For this reason, we used the *ratio* between reported literary and non-literary reading hours as a means of comparing our subjects. This ratio should, at the very least, reflect the subject's subjective sense of how much time he or she devoted to reading literary writing relative to non-literary writing, and hopefully separates out fiction and poetry readers from magazine and front-page readers.

In future studies, it may be possible to use more "objective" measures of reading habits— for instance, by examining the number of literary and non-literary authors each subject can identify (Mar et al., 2006), or by having subjects track their reading habits over time. Alternatively, we might conduct a study in which we ask a certain group of subjects to exclusively read literary texts for an extended period, while having another group read exclusively non-literary texts, and then measure the effect.

Our preliminary findings on the subject suggest that each person's model of the language they speak may be affected and "trained" over time by the specific linguistic samples they encounter. Intriguingly, differences in these individual language models appear to correspond with differences in "subjective" perceptions and judgment. Here, we examined how prior reading exposure may affect our perception and judgments of reading new texts. If our findings generalize to different genres of writing, spoken language, or even other modes of art and communication, we may be able to begin to explain individual differences in judgment and perception, and also how one can acquire taste through experience.

Acknowledgments

This material is based on work supported by the National Science Foundation under Grant Nos. 0547775 and 0624345 to Michael Ramscar, and by the Stanford University Undergraduate Advising and Research under Small Grant No. 4174 to Justine Kao.

References

Altmann, G.T.M. & Steedman, M. (1988) Interaction with context during human sentence processing. *Cognition*

Anderson, R., Wilson, P., & Fielding, L. (1988). Growth in reading and how children spend their time outside of school. Reading Research Quarterly

Baayen, R. H. (2001). Quantitative aspects of morphological productivity. In G. E. Booij and J. van Marle (eds), Yearbook of Morphology 1991, Kluwer Academic Publishers, Dordrecht

Bell, A., Brenier, J., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60(1)

Biber, D. (1988). *Variation across Speech and Writing*. Cambridge, UK: Cambridge University Press.

Biber, D. (1993). Representativeness in Corpus Design. *Literary and Linguistic Computing* 8 (4)

Bod, R., J. Hay, & Jannedy, S. Eds. (2003). *Probabilistic linguistics*. Cambridge, MA, MIT Press

Bybee, J. and P. Hopper, Eds. (2001). *Frequency and the emergence of linguistic structure*. Typological studies in language, vol. 45. Amsterdam, Netherlands, John Benjamins Publishing Company.

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

Bybee, J. (2006). From usage to grammar: the mind's response to repetition. *Language*

Cipielewski, J., & Stanovich, K.E. (1992). Predicting growth in reading ability from children's exposure to print. *Journal of Experimental Child Psychology,*

Cover, T.M. & King, R.C. (1978). A Convergent Gambling Estimate of the Entropy of English. *IEEE Transactions on Information Theory*

Davies, Mark. (2008-) The Corpus of Contemporary American English (COCA): 400+ million words, 1990-present. Available online at http://www.americancorpus.org.

DeLong K. A., Urbach,T.P., Kutas, M. (2005) Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*

Dixon, P., Bortolussi, M., Twilley, L. C. and Leung, A. (1993) 'Literary Processing and Interpretation: Towards Empirical Foundations', *Poetics*

Eisenbeis, R., and R. Avery (1972). *Discriminant Analysis and Classification Procedures: Theory and Applications*. Lexington, Mass.: D.C. Health and Co.

Hale, J. (2003). The Information Conveyed by Words in Sentences. *Journal of Psycholinguistic Research*

Karlgren, J., and D. Cutting (1994). Recognizing text Genres with Simple Metrics Using Discriminant Analysis. In *Proc. of the 15 '1' International Conference on Computational Linguistics (COLING '94)*.

Lee, Y.B. & Mayeng, S.H. (2002). Text Genre Classification with Genre-Revealing and Subject-Revealing Features. *Proceedings of the 25th ACM SIGIR Conference*.

Levy, R (2008). Expectation-based syntactic comprehension. *Cognition*

Mar, R.A., Oatley, K., Hirsh, J., dela Paz, J., & Peterson, J.B. (2006). Bookworms versus nerds: Exposure to fiction versus non-fiction, divergent associations with social ability, and the simulation of fictional social worlds. *Journal of Research in Personality*

Miall, D.S. & Kuiken, D. (1994). Beyond Text Theory: Understanding Literary Response. *Discourse Processes*

Mukarovský, J. (1964). Standard language and poetic language. *In P. L. Garvin (Ed.), A Prague School reader on esthetics, literary structure, and style* Washington, DC: Georgetown University Press. (Original work published 1932.)

Otten, M., & Van Berkum, J. J. A. (2008). Discourse-based word anticipation during language processing: Prediction of priming? *Discourse Processes*

Pierrehumbert, J. (2001) Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee and P. Hopper (eds.) *Frequency effects and the emergence of lexical structure*. John Benjamins, Amsterdam.

Pierrehumbert, J. (2003) Probabilistic Phonology: Discrimination and Robustness. In R. Bod, J. Hay and S. Jannedy (eds.) *Probability Theory in Linguistics*. The MIT Press, Cambridge MA

Ramscar, M., Matlock, T., & Dye, M. (in press) Running down the clock: the role of expectation in our understanding of time and motion. *Language and Cognitive Processes*

Servan-Schreiber, E., & Anderson, J.R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*

Wicha, N. Y., Bates, E. A., Moreno, E. M., & Kutas, M. (2003). Potato not pope: Human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neuroscience Letters*

Xiao, Z.H. & McEnery, A. (2005). Two Approaches to Genre Analysis: Three Genres in Modern American English. *Journal of English Linguistics*