

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

Here we examine how reading habits affect judgments of writing quality by looking at how readers' experience with reading shapes their sensitivity to the distribution of words in different texts, and how this in turn may shape what they recognize as good writing. We manipulated literary and non-literary passages so that the modified versions had lower word chunk frequencies but higher individual word frequencies than the original, and asked subjects to rate the quality of writing in each passage. Our results indicated that subjects who read more literary writing rated original literary passages higher, while non-literary readers preferred modified versions. Literary and non-literary readers alike rated the original versions of non-literary passages higher. Over time, exposure to different linguistic models appears to affect people's sensitivities to word and chunk frequencies, and may have implications for how a fine-tuned understanding of the probabilistic structure of language can be acquired through reading.

Keywords: Cognitive Science; Philosophy; Psychology; Communication; Language understanding; Skill acquisition and learning; Human experimentation; Literature

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 in it?

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; Cipelewski & 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 (Marr 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, and how this sensitivity may in turn shape judgments of writing quality. 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 work '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 higher frequency word than *task*, "a daunting task" appears over a hundred times more frequently on Google than "a daunting job." We are sensitive to the different frequencies of the two chunks and prefer the one with the higher frequency, even though there is no real reason why a job cannot be described as daunting.

The reason why we can sense the mismatch so easily 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 & Kuikan, 1994, Mukarovsky, 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 gets at the deeper point that the words in a given piece of writing will have different distributions depending on the corpus you are looking at (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 human readers. 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 in non-fiction than fiction), 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 looking at 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 and manipulated the frequencies of several chunks (short sequences of words) within each selection. We made the overall chunk frequencies lower, while actually raising individual word frequencies. We hope to examine whether subjects' evaluations of writing quality will reflect these changes, and whether the results will differ for literary and non-literary readers' evaluation of literary and non-literary texts.

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

In order to test our prediction that literary texts tend to contain more literary words and non-literary texts tend to contain more non-literary words, we ran each word in the passages through COCA to find its frequency in the fiction corpus, the newspaper corpus, and the general corpus as a whole. The average log frequencies of the passages in the three corpora are shown in figure 1. Our analysis revealed that within the specific corpora, the literary passages had significantly higher 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 passages in the fiction and newspaper corpora were not significantly different ($t(584) = -1.0288$; $p > 0.05$). This suggests that most words occurring in literary texts do not appear often in non-literary texts, while words in literary texts still occur fairly often in literary texts. Further, in the overall corpus, the literary passages also had higher average frequencies than the non-literary passages as well ($t(627) = 2.2786$; $p < 0.05$). 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) \times 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 ($F(1, 627) = 121.926$, $p < 0.001$).

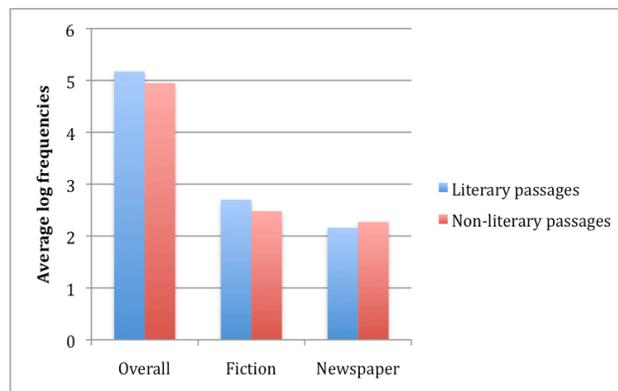


Figure 1. Average log frequencies of passages in different corpora

A more fine-grained analysis of the texts further supported the idea that the difference between literary and non-literary texts derives from the different ways that vocabulary is used in specific kinds of writing. We considered the pair-wise frequency of each word in each specific corpus. The average frequencies of the words in the literary passages were significantly higher in the corpus for fiction writing than in the corpus for newspaper writing ($t(335) = 11.4987$; $p < 0.001$), while the reverse was true for each of the words in non-fiction passages ($t(292) = -4.7295$; $p < 0.001$). In other words, what sets literary writing apart from non-literary writing is not that it makes use of a specialized set of literary words, but rather that it uses words in specialized ways that appear more frequently and perhaps selectively in literary writing.

Manipulation of Passages

After verifying this, we then manipulated the frequencies of three to seven chunks within each literary and non-literary passage, lowering chunk frequency while simultaneously raising individual word frequency. We measured individual word and chunk frequency based on the number of ‘hits’ the word or chunk had on Google. In the following example, (a) is the original chunk, and (b) is the modified chunk.

- (a) On the further *side* of the field¹
- (b) On the further *part* of the field

While *side* and *part* have highly similar meanings in this context, *side* has 805,000,000 hits on Google, while *part* has 1,750,000,000 hits, suggesting that *part* is a more frequent word in English than *side*. However, the original chunk (a) has 898 hits on Google, while the modified (b) has 0 hits. Although the average frequency of all words in (b) is higher, its chunk frequency is significantly lower than (a). What this suggests is that although *part* is a word that is used more frequently than *side* in general, it is not a word that is often used in this particular linguistic context (or chunk), even though it is similar in meaning and just as “grammatically acceptable.”

¹ Taken from “Emergency,” by Denis Johnson

Since we wished to manipulate word and chunk frequencies while keeping the meaning of the passages relatively constant, we selected chunks to modify based on whether or not they contained a word that could be replaced with a synonym that had a similar or higher frequency. While it is unclear if *all* the chunks we selected in our literary passages are used in exclusively literary contexts (e.g., “on the further side of the field” may be used in sports writing as well), we assumed, given the results of our passage assessment, that chunks from literary writing would be more familiar to literary readers.

Procedure and design

All surveys were designed and distributed using Qualtrics, the 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.

Participants were asked to take the survey individually on a computer. They were reminded to read the instructions in the survey carefully. We recorded the time it took each subject to complete the survey 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. After they 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 ratio we will refer to as the ‘literary reading score.’ We decided to 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 a salient feature of their reading habits.

Results

Based on their literary reading scores, subjects were divided equally into two groups 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. A repeated measures ANOVA of participant ratings of the modified and non-modified passages revealed a significant interaction between literary reading scores and within-genre preference ($F(1, 21) = 3.095$; $p < 0.05$). 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 ($t(29) = 1.7377$; $p < 0.05$). For non-literary passages, there was no significant difference between the within-genre preferences of subjects in the two groups, and they both preferred the original non-literary passages ($t(29) = 0.6556$; $p > 0.5$).

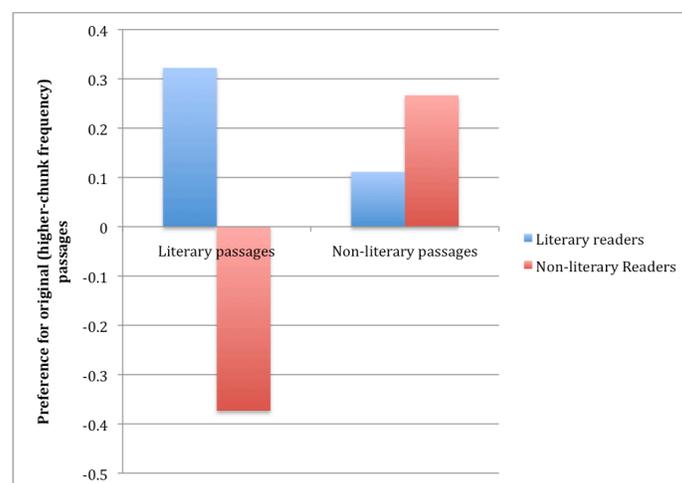


Figure 2. Literary and non-literary readers’ preference for original passages in the two genres

Discussion

As predicted, there was a significant interaction between subjects’ reading habits and their reading preferences. We hypothesized that readers’ tastes would be influenced by the presence of words (or rather, distributions of words) that were familiar to them. Our results seem to have borne this out. When reading literary texts, literary readers preferred the original passages, which were in keeping with the genre, whereas non-literary readers did not, preferring passages containing words that were likely more familiar to them. However, when reading non-literary texts, both literary and non-literary readers preferred the originals. These results indicate that literary texts require readers with a trained eye to detect subtle differences in chunk frequency, whereas readers do not need to read proportionally more non-literary writing in order to detect the same kinds of differences in non-literary texts.

How might one explain these results? Only people who are exposed to the distributional properties of those words in literary contexts appear to 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 our social and cultural lives naturally enforce a non-literary expertise on all readers, while literary expertise is more a matter of individual practice. Another reason may be that non-literary writing makes use of less specialized distributions. 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.

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 our fiction and poetry readers from our 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 has an individual model of the language they speak, which is trained over time on the particular set of linguistic samples to which they are exposed. 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 upon work supported by the National Science Foundation under Grant Nos. 0547775 and 0624345 to Michael Ramscar.

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 'I' 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*