effects, and how the term hacker once synonymous with innovation was effectively hijacked to denote cyberterrorism.

Chapter 5, 'Machinima and the Suspensions of Animation', deals with machinima (the production of animated films with video game graphics) and using imperfect, sometimes very low fidelity, appropriated gestures for conveying narrative.

At points, these chapters seem tenuously connected with the topic of the book, which is possibly symptomatic of the fact that the bulk of the book appears to be made up of a series of previously published essays. 'Glitch', 'noise', and 'error' are some of the most contextually sensitive terms that require a good deal more expansion than the book affords them. One could assume that Krapp is trying to make his readers work hard at finding the tangential nuggets of information amidst the noise and arriving at their own conclusions, but there really is no need for obfuscation, lack of clarity, or conciseness. Other reviewers (e.g. Leorke, 2012) have certainly taken more from this book and lavished it with a good deal of praise, but I feel they have been far too generous. In chapter 3, 'Noise Floor', Krapp states that 'the selection of a message in a set of possible messages to be observed in a noisy channel comes down to our acculturation to media technology' (p. 71). In this sense, the book itself is a bit noisy, as it assumes a certain deal of familiarity and acculturation to the field on the part of the reader for nearly all concepts being presented—not just media technology, but also the theoretical frameworks being discussed.

Put more precisely, there is not a highly memorable thread in Noise Channels that weaves all chapters together, or if there is one, it bears repeating, or struggles to manifest itself explicitly, specifically in relating chapters 1 ('Hypertext and its Anachronisms'), 2 ('Terror and Play'), and 5 ('Machinima and the Suspensions of Animation'). In contrast to chapters 3 and 4, these chapters seem highly tangential, and as such, the book drifts away from becoming the exhaustive and thorough examination of 'glitch', 'noise', and 'error' that the book's title would have you believe. For a clearer and less tangential discussion of the topics in chapters 3 and 4, I would highly recommend Rosa Menkman's 'The Glitch Moment(um)' (2011). Although this work, too, is composed of a series of texts written by the author between 2009 and 2011, they have been extended and reworked in a far more cohesive fashion.

References

Fuller, M. (ed.) (2008). *Software Studies: A Lexicon*. Cambridge (MA): The MIT Press.

Kahn, D. (2001). *Noise, Water, Meat: A History of Sound in the Arts*. Cambridge, MA: The MIT Press.

Kelly, C. (2009). *Cracked Media. The Sound of Malfunction*. Cambridge, MA: The MIT Press.

Leorke, D. (2012). Book Review: Noise Channels: Glitch and Error in Digital Culture. *Westminster Papers*, **9**(1): 171–178.

Menkman, R. (2011). *The Glitch Moment(um)*. Amsterdam: Institute of Network Cultures.

Nechvatal, J. (2011). *Immersion into Noise*. Ann Arbor: Open Humanities Press/MPublishing.

Iman Shayan Moradi
Running in the Halls Ltd, UK

doi:10.1093/llc/fqt004
Advance Access published on 20 February 2013

The Secret Life of Pronouns. What Our Words Say About Us. James Pennebaker. *New York: Bloomsbury Press, 2011. xii + 352 pp. ISBN 978-1-608194-80-3. $28.00 (hardback).*

This is a great book that aims to popularize the study of how function words such as pronouns, but also articles, prepositions, and auxiliary verbs, reveal personality traits and roles within relationships. James Pennebaker is a social psychologist who has made major contributions in understanding how people who have gone through traumatic experiences may be helped by writing about them. He has also invested a good deal of time in developing techniques for counting function words and interpreting differences in their distributions. And while Pennebaker focuses on interpreting differences in distributions as reflections of different personality types or different roles in relations,

he is interested in a wide range of other topics in which the interpretation of the word frequencies might play a role, including psychological health, emotions, honesty vs. deception, corporate and regional identity, literature, authorship attribution, authority in relationships, and political appeal.

The book deserves a review in LLC because it pays attention to linguistic and literary interests, and especially because it adds an interpretive dimension to *stylometry* (the study of style using exact techniques), which has been underdeveloped to-date (but see, too, Noecker *et al.*, 2013). As readers of this journal know, stylometry has come to focus increasingly on authorship attribution as an objective validation of its work, and has come to accept—with some demurring voices (Burrows, 2007)—that function word distributions are the most interesting indicators of authorship. I will criticize Pennebaker a bit later in the text for largely ignoring the stylometric literature, but I will focus on what he does contribute, and that is a great deal. The focused contribution Pennebaker is making concerns the interpretation of stylometric results—e.g. what does it mean if we find that an author uses an unusual number of first person singular pronouns (*I*-words)? His results are often surprising.

Over the years, Pennebaker has collaborated with a number of colleagues who he acknowledges generously throughout the book. A particularly important collaborator was Martha E. Francis, who developed the program Linguistic Inquiry and Word Count, which inputs document collections and outputs lists of word frequencies, normally classified according to semantic fields, such as anger, sadness, anxiety, or positive emotions. This seemed rewarding, but other collaborators soon convinced Pennebaker that there was more going on with function words, and these quickly became the focus of many years of inquiry.

For more than fifty years, work on non-traditional authorship attribution has focused on analysing distributions of function words. There are two reasons for this: first, because only function words occur frequently enough for reliable statistical inference, and second, because it is unlikely that authors attempt to manipulate their use of function words (Nerbonne, 2007). Stylometry is the more general effort to study textual style using exact techniques. Although there are excellent examples of stylometry in which, e.g. grammatical features are the focus of analysis (Baayen *et al.*, 1996; Hirst and Feiguina, 2007), and while there is an ongoing debate about how many common words to use, the dominant trend is definitely to use word frequency distributions of frequent words, i.e. function words. What is often missing, on the other hand, is a careful interpretation of what differences in function word distributions mean. Hugh Craig (1999, p. 103) asked the question most pointedly: 'If you can tell authors apart, have you learned anything about them?' Pennebaker's work is well-poised to fill in that interpretive gap.

For example, early in his book (chapter 3), Pennebaker notes that women tend to use more personal pronouns, negations (e.g. *no, not,* and *never*), 'certainty words' such as *always* or *absolutely*, and hedge phrases (*I think*). As a social psychologist, he elaborates on what this reveals about sex differences, even examining the lexical patterns of people undergoing sex-change operations, and testing whether masculine patterns correlate with varying testosterone levels, as the people involved received hormone injections. Pennebaker also examines correlations of function word distributions with age (younger writers use the past tense more, older writers the future tense) and class (upper classes use *we*, whereas lower classes emphasize *I*). In every case, Pennebaker proceeds from speech or text from people whose properties he can gauge, so when he characterizes women's speech as involving more personal pronouns, he can back this up with statistics compiled from empirical data. It is exactly this empirically validated level of interpretation that we normally lack in stylometry.

But, intriguingly, Pennebaker turns almost immediately to the question of whether (and which) authors tend to portray men and women faithfully in how function words are used. Joan Tewkesbury and Thorton Wilder do well in portraying men and women with respect to their distinct uses of function words, whereas Nora Ephron's and Woody Allen's characters all have feminine

distributions, and Quentin Tarantino's and William Shakespeare's characters tend to show masculine distributions—including the female lead in Romeo and Juliet! Naturally, one can argue with Pennebaker's characterization. For example, he does not mention verifying that the differences in the function word distributions he examines were the same during Shakespeare's time. But that is a quibble, compared with the enormous leap in interpretation Pennebaker is facilitating. Pennebaker and Ireland (2011) pursue this line of research in more detail.

In a chapter on detecting emotions in function words, Pennebaker not only introduces the different expressions of emotion using poetry, he goes on to devote sections to the language of suicidal poets (heavy on so-called *I*-words), and to the changes in King Lear's language as he came to realize the earlier errors he had arrogantly made. In a chapter on the language of lying, a major focus is the language of the self-deceptive Ebenezer Scrooge in Dickens' *A Christmas Carol*. In a chapter on how style indicators for partners in relationships tend to approach each other, he again turns to literature, examining the poetry of Elizabeth Barrett and Robert Browning on the one hand, and that of Sylvia Plath and Ted Hughes on the other. But in contrast to most stylometric studies, Pennebaker regularly backs up the claims he makes about writers and literary characters with empirical studies of the language of depression, deception, and strong emotion. His stylometry not only characterizes differences, it interprets them as well, and the interpretation is subject to empirical verification.

Pennebaker emphasizes a number of times in his book that he does not imagine that function word distributions *cause* the various correlates he studies (e.g. using positive emotion words does not improve mental health (p. 14), even though the use of these words correlates with better health); instead, the function words should be understood as indicators (p. 14, p. 102). He speculates that conversation partners are attuned to the signals the function words carry, even suggesting that mirror neurons might be involved (p. 202), and noting that adaptation to conversation partners happens quickly ('in a matter of seconds', p. 225). But function words do not signal individually but rather in the aggregate (as distributions), which means that they signal weakly. In one case, Pennebaker notes that Nixon used only 3.9% *I*-words in speaking with his aides, who used 5.4%. But for a signal involving a 1.5% difference in frequency to be interpreted reliably, one would have to be involved in minutes of conversation, not several seconds. (A back-of-the-envelope calculation suggests that the percentage difference 1.5% would translate to two standard errors after about 400 words, or about 2 min of rapid speech. The signal is just not available 'in seconds'.) This is probably unimportant for literary analysis, but it suggests that the psychological mechanisms influencing function word distributions may be more involved.

I have emphasized the real and potential contributions Pennebaker is making to stylometry, but I should not close without mentioning that the book is not without flaws. Some sections do not rise above banality, e.g. the first third of chapter 9 that explains how employees who refer to their departments, divisions, and companies using *we, our,* and *us* are more likely to identify with their employers and organizations. Stylometry experts may wonder how anyone could examine Jane Austen's vocabulary without so much as mentioning the Busa prize recipient, J. F. Burrows, but Pennebaker manages to compare works from different times in Austen's career (pp. 65–66) without so much as a nod to Burrows' (1987) chapter on 'The Changes that Time Brings'. In fact the only stylometry mentioned is Mosteller and Wallace's famous authorship study (whose work he takes care to 'correct', without mentioning any of the nearly 100 follow-up studies that Joe Rudman discussed at Digital Humanities 2012). And as one reads Pennebaker's forays *inter alia* into detecting psychological health, emotions, corporate and regional identity, authority in relationships, literature, authorship attribution, and political appeal—all on the basis of function word distributions, Maslow's (1966) remark is hard to repress: '[. . .] it is tempting, if the only tool you have is a hammer, to treat everything as if it were a nail'.

Acknowledgements

I am grateful to Karina van Dalen-Oskam, Maciej Eder, Jamie Pennebaker, and Jan Rybicki for discussing this review with me, naturally without attributing any of its views to them.

References

Baayen, H., van Halteren, H., and Tweedie, F. (1996). Outside the cave of shadows. Using syntactic annotation to enhance authorship attribution. *Literary and Linguistic Computing*, **11**(3): 121–31.

Burrows, J. F. (1987). *Computation into Criticism: A Study of Jane Austen's Novels and an Experiment in Method*. Oxford: Clarendon Press.

Burrows, J. F. (2007). All the way through: Authorship in different frequency strata. *Literary and Linguistic Computing*, **22**(1): 27–47.

Craig, H. (1999). Authorial attribution and computational stylistics: If you can tell authors apart, have you learned anything about them? *Literary and Linguistic Computing*, **14**(1): 103–13.

Hirst, G. and Feiguina, O. (2007). Bigrams of syntactic labels for authorship discrimination of short texts. *Literary and Linguistic Computing*, **22**(4): 405–19.

Maslow, A. (1966). *The Psychology of Science. A Reconnaissance*. New York: Harper and Row.

Mosteller, F. and Wallace, D. L. (1964) [2007]. *Applied Bayesian and Classical Inference: The Case of The Federalist Papers*. New York: Springer, 1984 [CSLI Publications published a reprint of the second edition in 2007 with a new foreword by John Nerbonne].

Nerbonne, J. (2007). The exact analysis of text. Foreword to the 3rd edition of Frederick Mosteller and David Wallace. *Inference and Disputed Authorship: The Federalist Papers.* CSLI: Stanford, xi–xx.

Noecker, J., Ryan, M., and Juola, P. (2013). Psychological profiling through textual analysis. *LLC: The Journal of Digital Humanities Scholarship* (available online Jan 8, 2013), doi:10.1093/llc/fqs070.

Pennebaker, J. W. and Ireland, M. E. (2011). Using literature to understand authors: The case for computerized text analysis. *Scientific Study of Literature*, **1**(1): 34–48.

John Nerbonne
University of Groningenm

doi:10.1093/llc/fqt006
Advance Access published on 6 February 2013

Corpus Linguistics: Method, Theory and Practice. Tony McEnery and Andrew Hardie. *Cambridge: Cambridge University Press, 2012. xv + 294 pp. ISBN 978-0-521-54736-9. £23.99, (paperback).*

Cambridge University Press already has three general books on corpus linguistics—Biber, Conrad and Reppen (1998), Meyer (2002), Hunston (2002)—so what does this new addition to the CUP catalogue provide that the others do not? As McEnery and Hardie explain in their preface, this book is not intended as an introduction to the practicalities of how to work with output from corpus analysis tools, such as concordance lines and lists of keywords. Instead, and here lies the book's distinctness, it concentrates on delivering rich discussions of fundamental conceptual issues and of the key trends in corpus linguistics, along with an account of how corpus linguistics has evolved and the major centres of development, with the main focus placed on English corpus linguistics.

The book is clearly a textbook—it has activities at the end of each chapter and a comprehensive (and lucid) glossary. The activities provided are stimulating and refreshingly generic (that is, they are not tied to particular worked examples, or to a single corpus or corpus analysis tool). There is also a website accompanying the book, which will provide answers to activities, extensions to some footnotes, updates to the book, and links to related websites; at the time of reading (17th January 2013), however, several parts of the website contained a message 'Coming soon!' last updated 31st October 2011. This lack of content is a shame because the concept and design of the companion website are good.

Because the book engages with theoretical and methodological issues rather than with worked examples, it is more appropriate for use with Masters-level students, or with advanced undergraduate students who are specializing in corpus linguistics. For universities with language/linguistics departments, this is an essential library book, but not necessarily a coursebook, for reasons, which should become clear below.

The book begins with a general introduction that addresses the question 'What is corpus linguistics?'