

- Gitelman, L.** (2006). *Always Already New. Media, History, and the Data of Culture*. Cambridge, London: MIT Press.
- Hoover, D. L.** (2004). Testing Burrows's Delta. *Literary and Linguistic Computing*, **19**(4): 453–75.
- Kirschenbaum, M. G.** (2006). *Mechanisms. New Media and the Forensic Imagination*. Cambridge, London: MIT Press.
- McGann, J.** (2001). *Radiant Textuality. Literature after the World Wide Web*. New York: Palgrave MacMillan.
- Miller, C. C. and Bosman, J.** (2011). E-Books Outsell Print Books at Amazon. *New York Times*, 19 May 2011. <http://www.nytimes.com/2011/05/20/technology/20amazon.html> (accessed 9 August 2011).
- Pierazzo, E.** (2011). A rationale of digital documentary editions. *Literary and Linguistic Computing*, **26**(4): 463–77.
- Rybicki, J. and Eder, M.** (2011). Deeper Delta across genres and languages: do we really need the most frequent words? *Literary and Linguistic Computing*, **26**(3): 315–21.
- Shillingsburg, P.** (2009). How Literary Works Exist: Convenient Scholarly Editions. *Digital Humanities Quarterly*, **3**(3): <http://www.digitalhumanities.org/dhq/vol/3/3/000054/000054.html> (accessed 9 August 2011).
- Stein, S. and Argamon, S.** (2006). A Mathematical Explanation of Burrows' Delta. *Digital Humanities 2006*. Paris: Sorbonne 5–9 July 2006: 81–82.

Thorsten Ries

Faculty of Literature and Philosophy, Ghent University, Belgium

doi:10.1093/llc/fqs006

Advance Access published on 22 February 2012

Ce qui compte. Méthodes statistiques. Ecrits choisis, tome II. Etienne Brunet (edited by Céline Poudat). Paris: Honoré Champion, 2011. 376 pp. ISBN: 978-2-7453-2225-8. €80 (hardback).

'Ce qui compte' ('what counts'), which should also be read as 'ceux qui comptent' ('those who count'), thus referring to statistical lexicographers, is the second volume of a compendium of papers by Etienne Brunet, a Professor at the University of

Nice who spent the last third of the twentieth century developing methods and tools in lexicometry.

For those who have never heard of Etienne Brunet, who essentially published in French, he could be depicted as the most prominent scholar of the French school of (computational) textual statistics, following the seminal role played in this respect by Charles Muller (1977), at a time when computers were not so common in the scholarly landscape. With a background in literary studies, he discovered the possible usage of computers for identifying the stylistic characteristics of a text in the course of his doctoral work on Jean Giroudoux and thereafter spent all his academic life exploring further the application of statistical methods to the understanding of lexical phenomena in texts.

Etienne Brunet entered the still to be explored field of textual statistics with a profound love for literary authors and their language. He appreciates each turn of language, knowing what is characteristic of the author; he can trace a phenomenon he observes within a weighted series of words down to the characteristics of a specific novel. You would hardly find a result which is not explained with reference to a general perspective on the author concerned (see his remarkable analyses on Hugo, Zola, or Proust). If I had to give a name to the discipline within which he should be situated, 'computational literary criticism' would definitely come to my mind.

With his deep understanding of the tenets of literary works, Etienne Brunet has developed incomparable expertise in the design of corpora, regarded as the experimental setting for his statistical exploration. In one of my favourite papers from the book (Brunet, 2004), actually situated in the context of a larger dispute on the authorship of plays attributed to Corneille and Molière, he gives a wonderful description of how he carved out a series of samples so that a colleague of his could carry out a blind experiment on the data. The fascinating analysis he makes of the criteria he chose and the impact of various possible sample combinations demonstrates how clearly he understood the effect of the corpus, his observable, on the results he could expect to observe.

A second essential skill that Etienne Brunet developed over the years was a deep and precise understanding of the behaviour of statistical measures in a text analysis context and their potential usefulness for specific inquiries. Among the statistical tools he learned to master, we should probably put to the fore all these having to do with the expression of contrasts within a corpus. One of the major aspects for which I will remember the work of Etienne Brunet is his contribution to the development of an in-depth methodology for contrastive analysis. For many years, his favourite method has been the use of standard deviation (*'écart réduit'*) as a general tool to observe differences within a corpus. Such differences may occur at many levels of analyses. At a macroscopic level, they can indicate the vocabulary characteristic of an author or a genre with respect to an appropriately varied corpus. They can elicit stylistic variations within an author's production across various periods of time for instance. This method allows one to analyse a given text in its general organization, looking for specific differences of expression within subsequences of the text. Among the most useful applications that Etienne Brunet makes of this analysis of difference is the observation of lexical development within a corpus subdivided into time-based subsets, as made available in the resources delivered with his software THIEF. It is also a tremendous tool to observe (*'sketch'*, in Adam Kilgariff's terms) words or expressions in context by the precise analysis of variations within a set of concordances, pointing out for example the impact of a morphological deference on a word form (for instance, a plural mark) in the semantic distribution of its co-lexicon.

Not only does Etienne Brunet like to observe the statistical behaviour of words, but he also takes a personal pleasure in observing the methodological effects of these tools on the legibility or reliability of his results. Moreover, he tends to be a purist in his defence of the simplicity of the Normal distribution, as opposed to the Poisson or the hyper-geometric distributions, in cases where the differences actually observed do not require the use of more complex models. His arguments are not purely mathematical; they have to do with the characteristics in size and

composition of his corpus and the nature of the feature he wants to observe.

It is important to situate this trend in contrastive analysis within the larger scholarly context in the France of his day. Etienne Brunet's work has a particular resonance for the linguistic community working on differential semantics, which, starting with Bernard Pottier (1974) and more recently under the leadership of François Rastier (1987), has advocated a description of linguistic phenomena based on observed contextual variations. Over the years, it is clear that a real community of practice has developed, sharing essentially the same vision of linguistic fundamentals.

In France, Etienne Brunet is also famous for producing software usually accompanied by *'textbases'* containing lexical data for one author or another. In 1989, he produced the first version of Hyperbase, a Windows-based program that is distributed along with an enormous amount of lexical and full-text data. The take-up of Hyperbase may have suffered from its proprietary implementation, as well as its absence from the OSS sphere (the software can be bought online from <http://www.unice.fr/bcl/spip.php?rubrique38>). As an illustration of this aspect of Etienne Brunet's production, I would encourage interested readers to have a look at and play with the online concordances and statistics he compiled concerning Balzac's *'Comédie humaine'* (Brunet and Kiriu, 1996). This is a kind of clear and simple presentation of lexicometric results that one would certainly show students wanting to understand the role of statistical methods in literary studies.

Finally, it is impossible to understand the work of Etienne Brunet, without noticing that he has been first and foremost a practitioner, who would systematically implement and experiment rather than develop new ideas on paper. Most of his writings reflect this: they are all essentially based upon graphic and other outputs produced by his own programs. Fully aware of this, he takes pride in his fondness for numbers at a period where most linguists regarded statistical methods with disdain:

Aux amateurs de grands espaces et de grands nombres s'ouvre l'aventure concrète et solitaire bien loin des théoriciens qui se

bousculent en piétinant la même surface pelée comme les minimes devant la cage de but’  
 [Those who love wide open spaces and large numbers are confronted by a well-defined and lonely adventure, far from the theoreticians scrabbling about in the same well-worn ground in front of the goal posts.]  
 (p. 45)

This citation provides me with an opportunity to comment on Brunet’s style, which in itself can be an obstacle to a full understanding of his contributions. It is almost as if, being constantly in contact with literary work, he needs to show that he too has a literary style of his own. His early work, in particular, is stylistically very creative—at times even borderline: ‘La statistique raffole des grands nombres comme les femmes faciles adorent les grandes fortunes’ [Statistics is crazy about large numbers, in the same way that easy women love the well-endowed] (p. 93)—if an article begins with a metaphor it will be drawn out as long as possible within the piece. At times, this can be really annoying (particularly for the non-native speaker) leading to impatience as the reader waits for the actual results of an analysis to appear. This is all the more unfortunate in that, when focusing simply on defending a point of view in the context of a scholarly dispute, he can actually write some of the most crystal-clear pages on statistical methodology available (see Brunet, 2004).

Putting all his energy within his software, Etienne Brunet also reflects his time with the absence of interest he shows in the way his data, whether text corpus or lexical results, could actually be made widely accessible by adopting standardized formats and devising a dissemination policy for his datasets independent of the software within which they are embedded. In his defence, he actually lived in a period when the largest database of French electronic texts, Frantext, was a bunker that could only be entered by a THIEF: the name he actually gave to software designed to ingest and manipulate Frantext data. Fortunately, times have changed. It has now become a basic tenet of research in digital humanities to think of the corpus as an object that should be published at an early stage.

Finally, the dated atmosphere that one can feel when reading the book is also reflected in the very existence of the volume itself. Written in French, this is a printed book, which only a few French-speaking university libraries will make available on their shelves. Is that really the best way to disseminate the work of Etienne Brunet to a new generation? Only a handful of these papers may be tracked down on the web, when, for instance, by happy accident they are part of the online proceedings of a conference, such as Brunet (1999, 2004, 2006a,b)—which I would actually recommend as starting material to those interested in the works of Etienne Brunet. Would it not have been much more forward-thinking to compile an electronic edition of these papers taking advantage of the secure environment offered by a publication archive—such as the French national service provided by the HAL platform: <http://hal.archives-ouvertes.fr/>? Perhaps translating some of the texts could have been too much of a challenge—especially considering the stylistic specificities of the author—but perhaps this may also be an important lesson for the young generation of French scholars. English is a tool that must be mastered if one is to make a long-lasting contribution to the world of scholarship, just as much as statistical methodology.

## Acknowledgements

With many thanks to Lou Burnard for proofreading the text.

## References

- Brunet, E. and Kiriū, K. (1996). *Balzac sur internet, Concordances et contextes de la Comédie humaine*. Base de données mise en place pour expérimentation le 21 avril 1996. <http://ancilla.unice.fr/~brunet/BALZAC/balzac.htm> (accessed 6 February 2012).
- Brunet, E. (1999). La base textuelle Batelier. In Delmonte, R. (ed.), *Proceedings of VEXTAL’99, Venezia, San Servolo V.I.U., 22-24 November 1999*. Venice: Venice University Press, pp. 175–83. <http://project.cgm.unive.it/events/papers/Brunet.pdf> (accessed 6 February 2012).

- Brunet, E.** (2004). Où l'on mesure la distance entre les distances. *Texto!*, IX(2): [http://www.revue-texto.net/Inedits/Brunet/Brunet\\_Distance.html](http://www.revue-texto.net/Inedits/Brunet/Brunet_Distance.html) (accessed 6 February 2012).
- Brunet, E.** (2006a). Le corpus conçu comme une boule. In Duteil-Mougel, C. and Foulquié, B. (eds), *Actes du Colloque international et école d'été 'Corpus en Lettres et Sciences sociales - Des documents numériques à l'interprétation', Albi, 10-14 juillet 2006*. Paris: Texto, pp. 69–78. <http://www.revue-texto.net/Parutions/Livres-E/Albi-2006/Brunet.pdf> (accessed 6 February 2012).
- Brunet, E.** (2006b). Navigation dans les rafales. *Actes JADT'2006, Besançon, Université Franche-Comté, 19-21 avril 2006*, pp. 15–29. [http://lexicometrica.univ-paris3.fr/jadt/JADT2006-PLENIERE/JADT2006\\_EB.pdf](http://lexicometrica.univ-paris3.fr/jadt/JADT2006-PLENIERE/JADT2006_EB.pdf) (accessed 6 February 2012).
- Muller, C.** (1977). *Principes et méthodes de statistique lexicale*. Paris: Hachette.
- Pottier, B.** (1974). *Linguistique générale*. Paris: Klincksieck.
- Rastier, F.** (1987). *Sémantique interprétative*. Paris: Presses universitaires de France.
- Laurent Romary  
Inria, France and Institut für deutsche Sprache und Linguistik,  
Humboldt-Universität zu Berlin, Germany  
doi:10.1093/llc/fqs008  
Advance Access published on 9 March 2012