Paola Vettorel[1]

INTERVIEW WITH MICHAEL RUNDELL – EDITOR-IN-CHIEF MACMILLAN LEARNER'S DICTIONARIES

April 8th, 2015

I interviewed Michael Rundell when he was invited for a series of Seminars at the University of Verona from March 25th to March 27th 2015. The Seminars were organized by the Dottorato in Lingue, Letterature e Culture Straniere Moderne, the Dottorato in Anglofonia and the Scuola di Dottorato in Scienze Umanistiche. Dr Rundell's seminars dealt with several aspects in corpora, from corpus-building and how to use corpora, to theoretical and methodological aspects of dictionary making, including most recent developments involved in the migration from print to digital media, crowdsourcing and "user-generated content".

Michael Rundell *has been a professional lexicographer since 1980. He has worked on several important dictionary and corpora projects, and is a founder member of EURALEX. In 2013 Coventry University awarded an honorary doctorate to M. Rundell "for his contribution to the description of the English language and to the field of pedagogical lexicography". He is currently Editor-in-Chief of the Macmillan learner's dictionaries and a trainer particularly in the areas of in lexicography and lexical computing. His publications include:*

Rundell, Michael. "The road to automated lexicography: an editor's viewpoint". Electronic Lexicography. Eds.Sylviane Granger, and Magali Paquot, M.Oxford: Oxford University Press, 2012. 15-30.
---. "It works in practice but will it work in theory? The uneasy relationship between lexicography and matters theoretical" (Hornby Lecture). Proceedings of the 15th EURALEX Congress. Eds. Ruth Vatvedt Fjeld and Julie Matilde Torjusen . Oslo: University of Oslo, 2012. 47-92
(with Sue Atkins). The Oxford Guide to Practical Lexicography, Oxford: Oxford University Press. 2008.

PAOLA VETTOREL: In the early 1980s you worked on the COBUILD project, which had important implications for language teaching, too (the *Collins COBUILD Dictionary for Learners*, the *Collins COBUILD Grammar,* and a COBUILD-based textbook). Can you tell us something about this groundbreaking project, and the kind of impact it had on ELT?

MICHAEL RUNDELL: COBUILD was an exciting place to be in the early 1980s. It was the first time a large corpus (by the standards of the time) had been used as a basis for creating dictionary text. All of us on the team, as we examined this new evidence of words in context, made new discoveries, almost daily, about word senses and especially about the way words combine to create meanings. The longer term effects were that the idea of a "lexical approach" to understanding how languages work - which scholars such as Harold Palmer and J.R. Firth had already glimpsed to some extent in the pre-computer age — became more fully articulated (above all by John Sinclair, the foremost linguist of his generation) and more solidly based on real evidence. This way of looking at language gradually trickled down into the language-teaching profession, replacing the older idea of grammar and vocabulary as distinct elements of the language system. I was only on the COBUILD project for about 18 months (before becoming Managing Editor at Longman Dictionaries in 1983) but it made a huge impression on me and on everyone else who worked there, and it led to the creation of the new field of corpus linguistics.

PV: In your career you worked at the development of several innovative learner's dictionaries, from the *Longman Language Activator* (1993) to the *Macmillan English Dictionary for Advanced Learners* (*MEDAL*,

---

[1] *Paola Vettorel (paola.vettorel@univr.it) is assistant professor of English Language and Linguistics in the Department of Foreign Languages and Literatures - University of Verona. Her recent publications include: (2013) "ELF in international school exchanges: stepping into the role of ELF users". JELF 2/1: 147-173; (2014) English as a Lingua Franca in Wider Networking. Blogging practices. Berlin: Mouton de Gruyter; (2015) (ed.) New Frontiers in Teaching and Learning English. Newcastle-upon-Tyne: Cambridge Scholars.*

1st edition 2002, 2nd edition 2007), that also won the Duke of Edinburgh/English Speaking Union English Language Book Award in 2002 and a British Council Innovation Award in 2004). In what ways were these learners' dictionaries new? What led to the need to re-think learner's dictionaries in new ways?

MR: The *Activator* was a completely different kind of dictionary (rather than an improved version of a regular learner's dictionary). It was designed to help users produce accurate, natural-sounding English in written or spoken mode. So it had two main goals. The first was to enable users to choose the "right" word - the one that was most appropriate for the situation. This meant that we had to use the evidence of our corpus (by that time, a 30-million-word collection) to clearly distinguish close synonyms: to explain, for example, the differences between clever, intelligent, bright, and brainy. This is not a strong point of conventional dictionaries, which focus more on distinguishing the different senses of polysemous words (what we call "word sense disambiguation"). So we had to devise checklists based on semantic, syntactic and sociolinguistic features (such as register and regional variety). The Activator's second goal was to enable a user, having chosen a word, to use that word appropriately and idiomatically, so there was a big effort to provide information on syntax, register, and phraseology. It was a great learning experience for all of us who worked on the project, and it's still one of my favourite dictionaries.

MEDAL came out in 2002, entering a crowded market which already included four good dictionaries aimed at advanced learners. However, we had a number of advantages because by the time we were developing MEDAL, access to large corpora and to good corpus-querying software had become much less problematic (and less expensive). Unlike some of the longer-established learner's dictionaries, MEDAL was compiled from scratch using corpus data, and it included three big new features, many of which have been taken on by other learner's dictionaries: a serious language-awareness component, which includes the popular "metaphor boxes" inspired by Lakoff and Johnson; a unique system for identifying a core vocabulary (of 7500 frequent words) which learners at this level should aspire to (these words are shown in red and divided into three frequency bands marked by stars); and a really systematic account of collocation, including 1000 "collocation boxes". This last feature benefited from the first use on a dictionary project of "Word Sketches", which have since become a standard tool for lexicography. This arose through a long-standing collaboration with my friend Adam Kilgarriff, a top computational linguist who (unlike most people in that community) has long experience of working with - and producing software tools for - lexicographers. There is a great demand for learner's dictionaries worldwide, and this created a busy market, which in itself has helped to drive innovation.

PV: An important aspect of dictionaries today is web-based resource tools,that have affected dictionaries in several ways; which are the most important aspects that have changed the making of dictionaries?

MR: In the area of dictionary development, things have changed dramatically in last 15 years or so - and in ways which make life easier for lexicographers, relieving them of many tedious tasks, and which also improve the quality of dictionaries. When we first had access to corpora, we had to scan concordances on printouts, and there wasn't always enough data to allow us to describe word behaviour with confidence. But today, corpora are far larger (for most European languages, corpora of a billion words or more are no longer unusual), and the software for extracting relevant information from them is very powerful. Equally, dedicated dictionary-writing software (which hardly existed 15 years ago) has improved working practices and brought a higher degree of consistency to what we do. In the UK at least, most of the dull routine tasks involved in compiling a dictionary have been transferred from humans to computers, and we continue to search for new ways of automating (or at least semi-automating) the various processes involved in creating new dictionaries.

PV: One implication of web-based dictionaries is the contributions users can make to them, e.g. to the Open Dictionary by MacMillan[2], that is, crowdsourcing and "user-generated content": what relevance does this have for dictionary-making today? What is the process that leads new words to enter (online) dictionaries in the 'digital age'?

---

[2] http://www.macmillandictionary.com/opendictionary/latestEntries.htm

MR: Crowdsourcing and user-generated content (UGC) are still quite new phenomena, and I don't think dictionary publishers have yet worked out how to maximise their potential. Our experience with the Open Dictionary is that it is good for picking up neologisms of all kinds, and for collecting "long tail" vocabulary, such as technical terms and items from the less dominant varieties of English. There is more to be done in terms of encouraging users to contribute, and providing clear guidelines and easy-to-use entry templates, so that more of what comes in is genuinely useful. A recent development with Macmillan's Open Dictionary is to integrate it more fully with the main dictionary; as an incentive to contribute, entries which started life in the Open Dictionary and were later "promoted" to become full entries in MEDAL will include a reference to the person who originally submitted the word and the date when they did it. (The word *selfie*, for example, originally came from a submission to the Open Dictionary, and has since been upgraded to become a full dictionary entry.) Crowdsourcing in its narrow sense refers to a model where a big task is "distributed" among large numbers of laypeople, and the data they create is then further processed by experts. A familiar example is the "reading programme" set up by James Murray in the 19th century. This involved thousands of volunteers collecting citations of words in use, which formed the raw material from which the Oxford English Dictionary (OED) was created. This approach, too, has plenty of potential for collecting language resources which may be useful for dictionaries, but things are still at an early stage.

As for words "getting into the dictionary", the migration of dictionaries from print to digital media has turned many of our old inclusion policies upside down. When dictionaries existed only in print, we had to develop clear and strict guidelines about what we could include, because space was always limited. With this constraint lifted, we could — theoretically - include anything and everything. In practice, we have had to devise new inclusion criteria relevant to the digital age. As always, we have to think about the needs of our main user group (people learning or using English as a second language), and it probably makes more sense now to think about why a word should *not* be included in the dictionary. For example, anything which is very parochial (of only local interest) is unlikely to be useful for our global readership; anything that looks like to be ephemeral (and will be forgotten in a few weeks or months) isn't going to be useful; and anything highly technical or specialised isn't appropriate for our audience. Apart from that, there is no good reason to exclude words for which there is good evidence of use. And again, developing these new criteria is work in progress, because it is only two years since Macmillan gave up printing dictionaries.

PV: English as a Lingua Franca is a particularly vibrant field of research that has led so far to the creation of three major corpora (VOICE, ELFA - WrELFA and ACE). Do you think that the actual use of English in LF settings can be taken into account in any way(s) in dictionary making, particularly as to language creativity (what Henry Widdowson refers to as the exploitation of the 'virtual language')?

MR: This is an interesting question. As a generalization, English dictionaries haven't paid too much attention to ELF thus far, on the grounds that dictionaries are there to describe the language as used by its native-speakers: this creates a standard to aspire to, without necessarily stigmatising other forms. But the picture is becoming more blurred as more and more interactions in English involve non-natives, and our corpus resources now include web data where some of the texts are produced by fluent speakers of English who have a different mother tongue. For example, there is growing evidence, as Anna Mauranen has shown, that some multiword expressions used as discourse organisers in academic writing started as ELF usages but have been taken on by ENL writers too, and this is the kind of thing lexicographers will need to pay attention to in the future.

PV: World Englishes and 'national' language varieties: how (and when) do nativised/local(ised) words and expressions enter a dictionary?

MR: Dictionaries of English have tended to focus on the two dominant varieties (American and British), but that is changing. Again thanks to the web, we now have much better language data to support improved coverage of "outer circle" varieties (and UGC can help here too, as mentioned above). In the next big update of the Macmillan Dictionary, which will go live in the autumn of 2015, we plan a major expansion of our coverage of the Englishes of South Africa, Australia, and India, with some material too from east Asian

locations such as Hong Kong and the Philippines. We've been working with colleagues in all these areas to collect new data, and there will be several hundred additions to the dictionary, which we hope will provide our users with the best possible coverage of World Englishes. But there are plenty of other parts of the world which we haven't yet focussed on, so - as always - there is more to do.

PV: How have online dictionaries impacted foreign language teaching and learning? Have they opened up new opportunities and approaches to dictionary use in the language classroom? And outside the classroom?

MR: That's not an easy one to answer. As the use of paper dictionaries declines (especially among the younger learners who form a major part of our user-group), we are seeing a rise in the use not only of laptops but especially of mobile devices (phones and tablets). And over half the people who arrive at the Macmillan Dictionary site come directly from a search engine like Google: they just ask for a definition or synonym, without necessarily specifying that they want a particular dictionary such as Macmillan. How this affects classroom activities is outside my area of expertise, but in a related field, the whole business of dictionary user-research is reinventing itself: traditionally, user-research has looked at how people interact with paper dictionaries, but in the last two or three years attention is turning to digital reference resources. Consequently, all this is still at an early stage, and the literature on using online dictionaries is not yet very extensive - but that will change. I seem to keep repeating myself here, but it is important to keep in mind that we are just at the beginning of a very big new paradigm for dictionaries. Printed dictionaries have been around for hundreds of years, but digital-only versions are still quite new, so we are working out the rules as we go along. It's an exciting time.

I would like to thank Michael Rundell for giving us his time for this interview.

Works Cited

Lakoff, George and Mark Johnson. *Metaphors we Live by*. Chicago: The University of Chicago Press, 1980.
Mauranen, Anna. *Exploring ELF: Academic English shaped by non-native Speakers*. Cambridge: Cambridge University Press, 2012.
Widdowson, Henry. *Defining Issues in English Language Teaching*. Oxford: Oxford University Press, 2003.