Everything You Always Wanted to Know about Dictionaries (But Were Afraid to Ask): A Massive Open Online Course

Sharon Creese[1], Barbara McGillivray[2], Hilary Nesi[1], Michael Rundell[3,4], Katalin Sule[4]

[1]Coventry University, [2]The Alan Turing Institute/University of Cambridge, [3]Lexical Computing, [4]Macmillan

E-mail: creeses@uni.coventry.ac.uk, bm517@cam.ac.uk, h.nesi@coventry.ac.uk, michael.rundell@lexmasterclass.com, k.sule@macmillaneducation.com

Abstract

We have created a Massive Open Online Course (MOOC) about dictionaries and dictionary-making, to be hosted by FutureLearn. This paper discusses the design and development of this course, which is pitched at high school and undergraduate level participants as well as language enthusiasts around the world. The MOOC will answer questions such as: how dictionaries are made and how this process has changed over time; what goes into a dictionary and who decides; and what kinds of language evidence underpin the information which dictionaries provide. Participants will be encouraged to compare the quantity and quality of information in different types of dictionary, and will investigate corpus-based and computational lexicographic methods. It will also consider dictionary users' attitudes and common misconceptions, taking into account the requirements and habits of English language learners as well as fluent speakers. By the end of the course, participants will know about some of the latest trends in lexicographic research, the roles of language technology, corpora and crowdsourcing in the dictionary compilation process, the range of possible dictionary entry components, lexicographical choices and computational methods surrounding the selection and ordering of word meanings, and the content and wording of definitions.

Keywords: MOOC, massive online open course, dictionary skills, lexicography, history of dictionaries, dictionary-making, corpus linguistics, neologisms, dictionary inclusion criteria, dictionary typologies, lexicographic evidence, crowdsourcing, meaning and definition, corpus-based lexicography

1   Introduction

This paper describes the design and development of a Massive Open Online Course (MOOC) on dictionaries, provided by FutureLearn and produced by Coventry University in collaboration with the Alan Turing Institute and Macmillan Dictionaries. The MOOC aims to provide an introduction to the world of dictionaries to a broad, non-technical audience, which includes language teachers and students but also laypeople with an interest in dictionaries and language. It is due to run for the first time in the autumn of 2018, and to the best of our knowledge it is the first MOOC in the field of lexicography.

FutureLearn is a private company owned by The Open University which provides a platform for a variety of courses created at other (mostly UK) universities and institutions. FutureLearn courses generally run from between two to eight weeks and are repeated at regular intervals. Participants enroll for free, although in most cases they can pay a small fee to retain access to the course materials after the course has ended. Institutions hope to gain a reputational advantage from their investment in FutureLearn course development, for example by establishing a positive relationship with users and attracting them onto related fee-paying programs offering certification.

FutureLearn takes a constructionist approach to knowledge acquisition, as described by Sharples (2018) and Sharples et al. (2012), and is based on the view that all human learning involves interaction. FutureLearn courses therefore restrict teacher input while maximizing the opportunities for peer-to-peer conversation. Source materials are generally expected to be provocative and short, and are always followed by at least one task requiring learners to post their own responses and comment on contributions from other learners. Video and audio recordings typically last no longer than four minutes, and more often are less than two minutes long. Written 'articles' are generally only a few paragraphs in length, providing a summary of the key ideas under discussion; links can be made to more extensive open-access material, but learners are not expected to have the skills to negotiate longer texts unaided: further reading activities are usually heavily guided.

MOOC tasks generally fall into two categories; feedback is either provided by peers (a model suited to the sharing of experiences and opinions), or automatically in response to quiz questions (a model more suited to the acquisition of factual information). Participants can pay to receive a certificate if they have completed the course, but they are not graded in any way, and do not produce assignments that require evaluation by a subject expert. This approach suits FutureLearn MOOC environments, where the number of online learners vastly exceeds the number of available moderators; it works best when the focal learning outcome is either an alteration in learner attitudes or the acquisition of factual knowledge, neither of which require scaffolding from experienced practitioners over an extended period of time. However FutureLearn MOOCs can be embedded within more conventional educational programs, such as face-to-face degree modules, and can thus be used to supplement and expand the learning experiences of students working towards the acquisition of more complex skills. Our dictionary MOOC might be incorporated not only into linguistics or lexicography courses, but also into courses in many other disciplines where an understanding of dictionary types and contents would help learners to communicate and study more effectively.

FutureLearn courses are divided into 'Weeks', described as "personally meaningful study periods" (Sharples, 2017); the MOOC we describe here consists of six Weeks, each providing about four hours of study time. The content of each Week is generally characterized by a 'big question', designed to stimulate participants' interest, and it is broken up into several topics or 'Activities', each with a defined goal. Within each Activity there are a series of learning tasks or 'Steps', each lasting about 20 minutes, which lead the way to the designated learning outcome. Activities are often introduced in a provocative way, and the first Step in an Activity is typically an invitation to discuss the topic. Reading or listening to input comes after this initial collaborative task, and is always followed by discussion and reflection. Later Steps in the Activity may include problem-solving, search engine browsing, investigating materials or authentic situations and co-creating or sharing artefacts. Learners on our MOOC will be asked to explore dictionary sites, for example, and create and share their own lexicographical material. Introductory textbooks such as Mugglestone (2011) cover some of the same ground, but the interactivity and interconnectivity of the MOOC makes for a rather different learning experience, adaptable to the learners' own dictionary-using habits and contexts.

Participants work at their own pace and can access any part of the course at any point, so the MOOC learning environment changes constantly as more and more contributions are submitted. One or more moderators see to the day-to-day management of these contributions, although on courses with high participation rates they cannot be expected to read all the comments that users post. The Week's activities are generally concluded with a round-up Step, asking participants to discuss what they think they have learnt so far. The course educators may also contribute summary postings at key points in the course.

Our MOOC is emphatically not a training course in dictionary-making or in the theoretical and computational research which underpins contemporary lexicography. Rather, it is intended to provide a

lively introduction to a topic about which non-experts often have strong opinions but little real knowledge. The course aims to bring users up to speed with recent developments in this rapidly-changing field. It will also challenge common misconceptions about dictionaries, such as the notion that it is a dictionary's role to pronounce on "correct" usage, or that a word's inclusion in "The Dictionary" confers some kind of official approbation. As well as explaining the kinds of information that dictionaries contain (much of which is often ignored by dictionary users), we focus especially on the evidence base of dictionaries, on issues around meaning and how it is encoded and understood, and on the technologies which are transforming dictionary-making and dictionary-publishing.

2   Week 1: Why Use Dictionaries?

We anticipate that the MOOC will attract a wide variety of participants from different educational backgrounds, and with different levels of linguistic and lexicographical expertise. The first Week of the course therefore starts by asking "Why use dictionaries when you can use search engines?", a question designed to provoke more conventional, and possibly older, dictionary-using participants, and also to acknowledge the reservations about dictionaries that other, possibly younger, participants might feel. Questionnaire findings suggest that learners do not always choose to use highly-regarded monolingual dictionaries, even if they consider these to be the best sources of information about words. We hope our 'big question' will indicate to participants that the MOOC is a non-judgmental environment, and will encourage honest reflection on possibly contradictory behavior and attitudes. The question may also expose differences of opinion regarding what actually constitutes a 'dictionary' and what constitutes a 'search engine'. The distinction might be puzzling to users if their search engine queries about word meaning or translation lead directly to online dictionary entries, without any need to specify a particular type of dictionary or to include mention of dictionaries in the query. Further confusion might be caused by the fact that, in some regions of the world, bilingual electronic dictionaries are not considered as 'dictionaries' at all, but are described simply as 'translators'.

The first Steps serve to introduce learners to the course educators and moderators and to each other. Following on from this, a series of Steps encourage consideration of possible reasons for dictionary use, as suggested in interviews with a representative selection of users, and in summaries of findings from various user surveys, from Barnhart (1962) and Quirk (1975), through to the recent large scale European survey on dictionary use. A central aim of the MOOC is to alter the participants' perceptions of the role of dictionaries in society. A poll will be taken in Week 1 to ascertain attitudes at the start of the course, and again in Week 6, so that participants can see if their own and others' attitudes have changed. The authority of dictionaries, differences between 'landmark' dictionaries such as the *Oxford English Dictionary* and less conventional products such the *Urban Dictionary*, and the rise of collaborative and crowdsourced dictionaries are all topics introduced in Week 1, to prepare the way for more in-depth investigations of dictionary content in later Weeks. We anticipate that as participants develop their ideas about the authority of different types of dictionary they will become more critically aware, and better able to distinguish between good and bad lexicographical practices.

3   Week 2: What's in a Dictionary Entry?

Week 2 examines the dictionary entry, and considers what kinds of information dictionaries need to provide for different types of word, and for different types of users engaging in different types of task. It begins provocatively, by suggesting that information encoded within an entry might be difficult to understand, and might often be ignored. Once again, this critical attitude towards well-respected

dictionaries is likely to surprise certain more traditional participants, but is also likely to reassure those who have themselves had difficulty interpreting dictionary entries; it should remind them that all kinds of opinion, including criticism of established practices, will add value to the discussion.

The tasks for Week 2 introduce participants to the basic components of the dictionary entry – the information that enables users to understand a word's meaning and use it appropriately. Participants will look at standard defining practices and their alternatives, which might involve illustrations, sounds, and perhaps even smells!. They will also be asked to consider the challenges of creating entries for words whose usage is restricted in some way, for example in terms of register, region, prosody or phraseology. Having broadened their understanding of the range of information that might be included, participants will examine specific dictionary entries, and consider the appropriacy of the information provided, and its accessibility to the user.

Of course dictionary users' information needs depend on the tasks they are engaged in at the time of dictionary consultation, so Week 2 includes consideration of the contexts of dictionary use, bearing in mind the portability of mobile phones and recent technological advances in speech production and voice recognition. Participants will be invited to share their experiences of dictionary use in both conventional and less-conventional settings (for example when speaking to shop assistants, or asking for directions on the street). A discussion of the value of different types of dictionary entry for different types of activity should also activate opinions concerning monolingual versus bilingual dictionaries, print versus e-dictionaries, and learner dictionaries versus those intended for fluent users of the language.

Thus, as well as introducing participants to some basic concepts, to be explored in greater depth in Week 5, the work in Week 2 helps to develop participants' critical faculties, and may challenge some of their long-held views about the authority of dictionaries.

4   Week 3: Evidence and Method: Where Does the Information in Dictionaries Come From?

Week 3 is dedicated to illustrating the evidence and the methods used by lexicographers. This is intended as an introduction to the dictionary-making process for non-specialists, as we believe that learning about evidence sources for lexicography can raise the participants' proficiency as dictionary users too. This Week builds on the previous ones, particularly the introduction to the content of dictionary entries in Week 2, and prepares the ground for the following weeks, particularly the dictionary inclusion criteria described in Week 4 and the definition-writing process explained in Week 5.

We encourage participants to reflect on the sources that could be relied on to create dictionary entries, before introducing them to an overview of evidence sources in lexicography. We then provide a historical overview of the use of evidence sources in lexicography, starting from pre-computational approaches based on paper slips, with reference to the *Oxford English Dictionary*. Following on from this we explain the "corpus revolution" and show how it brought fundamental changes to the dictionary-making process. Participants' engagement is facilitated by a series of tasks aimed at raising their awareness of the role played by corpora in lexicography, and a deeper understanding of the nature of dictionary content and its relation to linguistic evidence. The tasks include a comparison between a dictionary entry and concordances of the headword in a corpus, with the aim of discovering gaps in the dictionary entry, as well as exercises about using corpus data to create a dictionary entry and identifying the corpus evidence relative to different components of dictionary entries. All this is aimed at triggering a critical discussion of the relationship between dictionaries and language data, and appreciating some of the subtle points regarding admissible and suitable evidence in lexicography. This

way, the course participants will reach a fuller understanding of the nature of dictionary content and its relation to language.

We also cover the various stages in corpus-based lexicography, from corpus design and text collection, to corpus annotation and 'word sketches', so that participants see how linguistic analysis levels, such as morphological or syntactic and so on, can be surfaced in the lexicographic process, become familiar with such a critical aspect of dictionary-making practice, and recognize the challenges of dictionary-making. We finish by presenting the role played by corpora and quantitative information in understanding language and therefore describing it in dictionaries, and a comparison of how corpus evidence is reflected in different dictionaries. This will relate to the participants' personal experience with dictionaries and encourage them to be more aware users of dictionaries.

5   Week 4: What Goes into a Dictionary - Who Decides, and How?

Week 4 of the MOOC invites participants to explore what goes into a dictionary, who decides what is included, and how those decisions are made. This is done through a focus on the selection processes followed by dictionaries of various types (both 'traditional' dictionaries such as bilingual, monolingual and learner, and 'less-conventional' ones such as *Wiktionary* and the *Urban Dictionary*). Although inclusion criteria are more relaxed for less-conventional dictionaries than for their traditional counterparts, participants will be led to realize that attestation of real-world usage is crucial when differentiating genuine neologisms from 'buzzwords' (defined by Neuman, Nave and Dolev as 'fashion words that enter the language and rapidly acquire great popularity' (2010: 58, 67), yet regularly fade into obscurity).

We begin by asking learners to compare their estimates of the number of new words entering 'traditional' dictionaries each year with the figures given in the publicity material produced by renowned publishers of English dictionaries. From here participants are asked to consider why some words are accepted into dictionaries, while others are rejected. This will encourage discussion of the way some of the dictionary-making processes introduced in previous Weeks are put into practice, and will also introduce participants to some of the key issues facing lexicographers and collaborative dictionary contributors.

Participants are then asked to decide whether or not some newly-created words should be included in a dictionary of their choice, and to produce and discuss their own inclusion criteria, bringing their own practical knowledge and experience to bear. To support this discussion, we provide dictionary inclusion criteria from published sources, and input from professional lexicographers regarding the factors which actually influence the acceptance and rejection of new words or new meanings in a dictionary. The activity takes into account the different needs of different types of user: the *Oxford English Dictionary*, for example, focuses on those words deemed most likely to survive long term, rather than those which are most current (Algeo 1993: 283), while entries into *Wiktionary* tend to be more fluid. Of course, neologisms are not the only words added to dictionaries, so participants will also be introduced to other types of candidate, including borrowings, and words which have undergone a change of meaning or word class.

Finally, the Activities in Week 4 show participants how the lines between 'traditional' and 'less-conventional' dictionaries are becoming blurred, and how the use of crowdsourcing and user-generated content can add value to the former by incorporating elements of the latter. The concept of true 'crowdsourcing' and its application to dictionary-making is explored, along with the distinction between crowdsourced and strictly 'collaborative' dictionaries, and participants are asked to consider what collaborative and/or crowdsourced dictionaries are available in their own home countries, which ones they use, and why.

6   Week 5: Meanings and Definitions

Week 5 is perhaps the most intellectually-challenging part of the course. In this section, we tackle the subject of meaning: how meanings are created, why some words have more than one meaning, how lexicographers identify meanings, and how meanings are explained in a dictionary.

One of the big lessons of corpus linguistics is that the neatly-divided numbered "senses" in dictionaries imply a level of certainty around meaning which is not always supported by the evidence of language in use. So a major objective in this part of the course is to encourage participants to confront the reality that meanings are often less stable and less discrete than dictionaries suggest. As in much else of this MOOC, the message we want participants to absorb is that these things are not fixed "from above", but often require difficult judgement calls on the part of lexicographers.

We start by asking the question "How do we know what words mean?", and this is developed through a number of tasks, short articles, and videos. In one exercise, participants are presented with a series of (corpus-derived) sentences illustrating several uses of a polysemous word. Their task is to assign each sentence to a specific numbered sense in a dictionary entry, and it quickly becomes clear that such mappings are not always straightforward. In another Step, participants are introduced to the concept of polysemy, and they are asked to explain why this does not (as might be expected) lead to ambiguity and confusion among speakers and listeners (or writers and readers). We demonstrate that, in normal communicative interactions, fluent users of a language reliably identify the "right" intended meaning (when several possibilities exist), and we ask how they do this. Participants analyze corpus data to match a given instance with a specific meaning, and are then asked to find the contextual clues (syntactic, collocational, phraseological, and so on) which led them to associate each occurrence of a word with its dictionary meaning. This leads participants to the discovery that context almost always resolves any potential ambiguity. Following this, participants build up an inventory of criteria for distinguishing the various meanings of a complex word, and thus gain an understanding of how lexicographers approach the task of word sense disambiguation. This part of Week 5 ends with an article summarizing what has been learned about the relationship between meanings in the real world and "senses" in a dictionary.

In the second part of Week 5 we move on to the topic of definition, and think about how dictionaries explain meaning. Participants are first encouraged to give their own definitions of familiar objects and concepts, and then to share and discuss their output. This gives them an insight into the challenges involved, promotes an understanding that there may be several quite different ways of achieving the same result, and leads them to think about what definitions are actually for.

We then look at a range of dictionary definitions for the same "simple" word (such as the name of a familiar animal): why aren't these definitions all the same, and what factors account for the differences among them? This leads on to a discussion of the needs, prior knowledge, language proficiency and so on of different types of dictionary user, and how this affects the structure and content of definitions. This point is consolidated through a task where participants evaluate different definitions for the same word, and give their views on which is best (and what "best" means in this context).

The issue of what makes a good definition is further developed through a number of Steps (articles, videos, exercises). Looking at a range of definitions in well-known dictionaries, we focus first on the informational content required for a definition which will resolve a user's communicative problems: how much is enough, and is there such a thing as too much information? We introduce participants to the idea of the definition as a "typification", rather than an attempt to account for every conceivable use of a word in text. This leads to the recognition that the word "definition" itself is problematic: the Latin root implies a level of certainty and precision which is in unattainable for most items of everyday vocabulary.

Moving on to the language and structure of definitions, participants are introduced to a range of defining styles, from the most traditional ("the act of x-ing, etc."), through folk-defining techniques, to more contemporary approaches such as full-sentence definitions. The relative merits of these different styles are evaluated. We also address the issue of dictionaries' supposed objectivity – and its limits. The *Urban Dictionary* provides clear cases where a definition conveys opinions rather than facts, but although "serious" lexicographers strive to avoid subjective judgements when writing definitions, there are certain types of word where culture-specific norms are difficult to avoid – words such as *disabled, marriage, civilization*, or *god*.

In a final exercise, participants consolidate what they have learned by writing – and then discussing – their own definitions for selected words. A video in which the educators reflect on the main issues covered during the Week concludes this part of the course.

7   Week 6: What Does the Future Hold?

The final Week of the course is devoted to new research trends in computational lexicography (including automatic neologism detection and new sense detection) and dictionary use, as well as a summary of the course and a look at the future of dictionaries. The content is highly interactive, encouraging participants to reflect on their learning and how the course has changed their view of dictionaries. Participants will also be invited to discuss what the role of dictionaries in today's and tomorrow's world should and will be.

8   Conclusion

We have described what we believe to be the first ever MOOC devoted to dictionaries. We see this as a significant development, which will familiarize non-experts with the challenges faced by lexicographers and introduce them to the wide range of activity within our field – as well as, hopefully, dispelling some common misconceptions about dictionaries and how they are made.

The course will be offered free of charge, and is open to everyone, so some participants will probably choose to study it independently and on a voluntary basis. These might include language teachers and teacher trainees, and laypeople with an interest in language issues. However we believe that a significant number of participants will take the course as part of a credit-bearing program within their school or university, and we anticipate that some institutions will blend our online course content with input from their own teachers, to create modules tailored to the specific needs of their students. Clearly there is scope for learners to progress beyond the content of this MOOC to more advanced lexicography and dictionary user research. Once the MOOC is running, we will be able to monitor its effects on participants, and perhaps consider developing a second stage.

Although the course's primary purpose is pedagogical, the inclusion of tasks requiring participants to upload information about their own dictionaries, and (later) their own dictionary-using habits, could provide a rich source of research data, available not only to us, the course developers, but also to any participant who enrolls. If the MOOC is embedded within a formal university program, for example, this data could be used in student assignments, mini research projects, or even dissertations or theses, and the Terms and Conditions set by FutureLearn allow the anonymous data collected from the course to be used in this way.

In addition, we hope to gain valuable insights from participants' feedback and discussions throughout the course. This could well have value in informing dictionary content, developing teaching materials in relation to dictionary use, and as a data source for researchers at all levels.

References

Algeo, J. (1993) 'Desuetude among New English Words'. *International Journal of Lexicography* 6(4), 281-293.

Barnhart, C. (1962) Problems in Editing Commercial Monolingual Dictionaries. - In: F. Householder and S. Saporta (eds.) *Problems in Lexicography*. Publication twenty-one of the Indiana University Research Centre in Anthropology, Folklore and Linguistics, 161 - 181. Bloomington: Indiana University.

Kosem, I., Lew, R., Müller-Spitzer, C., Wolfer, S. (2017). The European survey of dictionary use. In *Electronic lexicography in the 21st century: lexicography from scratch, the eLex 2017 conference*, Leiden, the Netherlands (book of abstracts). Leiden: Dutch Language Institute; Brno: Lexical Computing; Ljubljana: Trojina Institute for Applied Slovene Studies, page 53.

Mugglestone, L.(2011) *Dictionaries: A Very Short Introduction.* Oxford: Oxford University Press

Neuman, Y., Nave, O. and Dolev, E. (2010) 'Buzzwords on Their Way to a Tipping Point: a View from the Blogosphere'. *Complexity*, 16(4), 58-68

Quirk, R. (1975): The social impact of dictionaries in the UK. In: R. McDavid and A. Duckert (eds.): *Lexicography in English*, 76 - 88. New York: Annals of the New York Academy of Sciences 211.

Sharples, M. (2018) *The Pedagogy of FutureLearn: How our learners learn*. Available at https://about.futurelearn.com/research-insights/pedagogy-futurelearn-learners-learn

Sharples, M. (2017) Pedagogy of FutureLearn: How our learners learn. Powerpoint presentation available at https://www.slideshare.net/sharplem/pedagogy-of-futurelearn

Sharples, M., McAndrew, P., Weller, M., Ferguson, R., FitzGerald, E., Hirst, T., Mor, Y., Gaved, M. and Whitelock, D. (2012). Theory and practice of teaching, learning and assessment. Innovating Pedagogy 2012: Open University Innovation Report 1. Milton Keynes: The Open University.

Acknowledgements