THE EFFECTS OF THE TRANSITION FROM PRINT TO DIGITAL DICTIONARY.

Karimbaeva Mavluda

Senior lecturer Andijan State University, Uzbekistan

Abstract: *Over hundreds of years, lexicographic conventions and editorial standards have been changed. They originated during a period when dictionaries were printed books with limited space, as they have been for nearly the entirety of their existence. Many times, the styles that we seem to be "natural" dictionaries' features are actually tricks meant to fit as much information as possible into the little amount of space that is available.*

A straightforward illustration is the type of "recursive" definition that can be found in many English dictionaries, where the user is required to perform a second look-up (to the base word) when a nominalization, like assimilation, is defined in terms of the related verb ("the act of assimilating or state of being assimilated"). Is this the best course of action, or was it chosen merely because it required less room than a self-sufficient explanation?

Space limitations vanish when dictionaries progressively move from print to digital medium. Some issues just go away. For example, tildes, abbreviations, and the like are no longer necessary, although surprisingly many dictionaries still adhere to these norms in their digital counterparts. Thus, the question of whether we should review and reassess the whole set of editorial guidelines and practices in light of the evolving situation emerges. This article examines various well-known presentational and editorial traditions, determining which ones are inappropriate for the digital era and proposing potential replacements in the form of new regulations.

Keywords: *definitions, printed volumes, example sentences, digital media, exclusion criteria, lexicographic conventions, online dictionary, user profile*

This article deals with a variety of well-known editorial practices and lexicographic conventions. The goal is to determine if regulations set during the lengthy era when dictionaries were exclusively available as printed books are still applicable in the twenty-first century, when many, if not all, dictionaries are now produced in digital medium.

Macmillan added entries for 64 chemical elements to their lexicon as part of a normal update in early 2015. This concluded the dictionary's coverage of all 118 elements. However, it is valid to question why they were not all included at the beginning. Dictionaries typically include all members of a specific set, such as days of the week or zodiac signs. However, in this case, rarer parts were omitted to focus on more frequently used vocabulary topics.

When dictionaries are published in the form of printed volumes, editors frequently make judgments like this: a book with finite dimensions creates a "zero-sum" game in which the addition of one category of information necessitates the deletion of another.

When major new editions are released (which happens every four or five years), the problem can worsen. Newly emerging words, phrases, and meanings must be contributed to guarantee that the dictionary remains current. To accommodate newcomers, we must decide whether to eliminate an equal quantity of material. What criteria should we use; whether to raise the size of the book (a popular solution, but unsustainable in the long run); or to generate additional space by making typographical changes and raising the amount of content on the page (which may alienate readers). Each method has its own hazards, which we normally aim to minimize by carefully calibrating a combination of all three expedients.

This is one of the reasons why digital media is so ideally suited as a platform for all forms of reference materials (dictionaries, maps, and encyclopedias).

In many regions of the world, paper dictionaries still have a bright future. Furthermore, certain types of dictionaries, such as those developed for schools, special-subject dictionaries, or dictionaries of "smaller" languages, may continue to prefer print for the foreseeable future. However, for three major categories of dictionary — which collectively account for a sizable portion of the global dictionary market and have had the greatest impact in terms of lexicographic innovation — the long-term decline in printed edition sales is irreversible, prompting publishers to focus more on digital versions.

There are three types of dictionaries: general monolingual dictionaries for adult mother tongue speakers, bilingual dictionaries for "big" language pairs, and monolingual learner dictionaries.

While there is currently some unequal progress, the direction of development is evident. Nor should this be interpreted as a reason for melancholy: new media, with their boundless space and digital features like multimedia and hyperlinking, offer exciting prospects for innovation and better coverage. They also create countless opportunities for reference materials that will better meet users' needs than their print-bound predecessors.

This article examines the effects of this shift on the kind and quantity of material included in dictionaries, as well as how they present it. It also assesses how successfully the current generation of online dictionaries has adapted to the new situation.

The printed book served as a major evolutionary platform for the dictionaries that we are familiar with today. This indicates that editorial guidelines and lexicographic norms have evolved and solidified over the course of more than 400 years in the English language.

People are aware of what to anticipate from their dictionary, as well as what not to expect: numbered word senses, succinct definitions using well-known (albeit occasionally unclear) formulas, methods for writing word sounds, and so forth. However, a lot of what

we consider to be "natural" dictionary properties are actually just conveniences. They developed in order to meet the necessity of cramming a lot of information into a little amount of space, not because they are the greatest means of informing people.

When the Macmillan Dictionary was first created for print, it was clearly designed to favor the most common terminology in the English language. The intention was to give a core collection of 7500 high-frequency terms comprehensive information (on syntax, collocation, phraseology, register, and so forth), supported by a wealth of sample sentences. This approach's inevitable drawback was that words outside of this group were treated more hurriedly and frequently had no examples at all.

While adding an example sentence at a word like parsimonious could mean that an important pattern at a verb like instruct would be left without an example, which could be problematic for the student who needs to use instruct productively, the policy is far from ideal but perfectly justified in the context of print publishing.

The definition of "dictionary" has already changed as a result of the digital revolution. Modern general-purpose monolingual dictionaries frequently incorporate one or more of the following: a blog, a thesaurus, multilingual content, games or puzzles pertaining to language, "Ask the Editor" features, videos, infographics, and various forms of user-generated content. The nearly continual activity on social media lends credibility to these. However, the primary purpose of a dictionary is to describe the meanings and usage of words in a language; the novel aspects that enrich and complement what is already there are not the focus of this study.

These include, in general, dictionaries that are only available digitally, those that are published in print, and dictionaries that are available in print and digital media. Even digital-only dictionaries are, for the most part, developed from print products, but the second category is arguably the most prevalent (as of this writing). One example of the latter kind is the Macmillan Dictionary, which debuted as a printed book in 2002. additionally, moving to a digital-only model in 2013.

Although this is an extreme example, there is often an uncomfortable mix of old and modern elements in online dictionaries. Despite never having been published in print, Wordnik's two-column layout displays definitions from a variety of conventional dictionaries on the left side of the screen, accompanied by examples of modern sentences taken from the web on the right. As a result, we can locate current instances of the social media sense at its twitter entry, but the definitions that correspond to it (which are based on an outdated version of the American Heritage Dictionary) do not capture this newly created meaning. One very fascinating example is Wiktionary.

This dictionary appears to be highly "modern" on the surface: it is fully web-based, with entries derived from user-generated content and no connection to conventional print lexicography. However, things are not quite that easy. Most definitions for more "everyday" vocabulary are just copies of definitions from other dictionaries, even though

the majority of entries for subject-specific terminology are newly written, usually by persons with specialized experience.

Even worse, authors of Wiktionary, who understandably worry about intellectual property rights, frequently appropriate passages from a 1913 version of Webster's Revised Unabridged Dictionary that is safely free of copyright violations.

The best "hybrid" dictionaries still contain vestiges of previous methods of operation (a dictionary written for print publishing is also available online). However, diligent work is being done to adjust to the new format. Dictionary creators are reconsidering the use of alphabetical order in addition to making obvious adjustments (such as writing out shortened forms and grammatical codes, adopting a more "open" style where different information kinds begin on a new line and frequently in a new color, etc.).

Alpha order is the method by which users in a conventional macrostructure locate what they're looking for. Early digital dictionaries stuck to this concept since it is so essential to the layout of print dictionaries: they kept listing dictionary entries alphabetically, not seeming to understand that this method was out of date for online resources.

Even before dictionaries transitioned from print to digital, publishers faced the problem of information overload since the corpus revolution made it possible to give more thorough explanations of a word's contextual aspects. Digital media helps publishers overcome these obstacles to some extent, though they are still figuring out the best ways to do so.

There is a limit to how much information can fit on a single screen, regardless of how much is in the dictionary database. Furthermore, the issue gets worse with the increasing tendency of using mobile devices to check dictionaries. However, dictionary creators also acknowledge that users consult dictionaries for diverse reasons in diverse contexts — generally speaking, in receptive or productive modes, but with different subcategories of these types of modes.

Because of the intersection of all these factors, publishers face the challenge of designing macrostructures that minimize the issue of "too much information" and make use of search strategies that web-savvy users are already familiar with, all the while facilitating access to various information layers that will meet a variety of use cases.

Thus, we are in a period of transition. The shift to digital media has brought forth both benefits and obstacles, which are generally well known. Publishers have realized that eliminating restrictions on "storage space" does not give them permission to compromise on the conventional values of conciseness.

As John Simpson has observed, "if editors were to allow the extent of individual entries to range out of proportion to utility this would result in making the user's task of interpreting an entry much more difficult" (Simpson 2014: 21). Though there's been a lot of experimenting, there's not yet been much agreement on how to proceed. Thus, now is an opportune time for a thorough reevaluation of the rules and practices that have grown

so ingrained that we could mistake them for being a necessary component of every dictionary.

What is needed now is "models for e-dictionaries that focus on critical areas like the data to be included … the structures to present and accommodate the data, the functions of these dictionaries and the way they should respond to the needs of their target users" (Gouws 2014: 157).

One of the first questions any dictionary publisher has to consider is "which words get into the dictionary". The theoretical background to this is the observation that the lexicon is an unbounded set. As Hanks points out, "the lexicon is dynamic: new words are being added all the time" (Hanks 2013: 29). Given that not even the powerful OED claims to contain every English term, it follows that all dictionary publishers must have strong inclusion policies. Do we need to reconsider these standards in light of the digital era?

The corpus evidence for a word's frequency, currency, and dispersion across text genres and geographies is usually taken into consideration when determining inclusion criteria.

Digital dictionaries are already modifying some of these requirements. For instance, the Oxford Dictionaries website discusses longevity and notes that it may not be suitable to use the older, more stringent standards for determining how long a word has been in use:

"It used to be the case that a new term had to be used over a period of two or three years before we could consider adding it to a print dictionary. In today's digital age, the situation has changed" (http://www.oxforddictionaries.com/words/how-do-new-words-enter-oxforddictionaries). However, the "user-profile" of a specific dictionary and the amount of space available are two important considerations.

Due to space constraints, dictionary creators must be selective about what they include, which has led to the public's perception of dictionaries as "gatekeepers," only allowing words that they approve into "the dictionary." The most trustworthy method of guaranteeing that the headword list that is produced is appropriate for its intended use is to have a strong user profile.

A user profile "seeks to characterize the typical user of the dictionary, and the uses to which the dictionary is likely to be put" (Atkins and Rundell 2008: 28). When space is restricted, having a good understanding of the target user's receptive and productive needs, prior knowledge, language competency, and reference skills is a crucial tool for inclusion decisions. However, when the dictionary is available online, neither factor is as important. Unrestricted space undoubtedly allows inclusion standards to be loosened, but it also makes it much more difficult to predict user identity in an online environment.

Regarding the well-known English monolingual learner's dictionaries, more than half of the users who visit the website did so via "organic search"—the user types a search term into their search engine (like "definition of X") without specifying a dictionary, and then clicks on one of the results. The "direct search" method, in which the user enters the

name of a specific source (like Oxford or Macmillan), makes up a lesser portion of all traffic to most dictionary websites.

As a result, it is more challenging to identify the possible user population, which increases the difficulty of feeling sure about inclusion decisions.

We need to examine a few particular categories after establishing a basic strategy based on "exclusion criteria". Among them, "named entities" are the most challenging; they are, in general, the names of individuals, locations, organizations, businesses, and so forth. Do dictionaries need to contain them? The majority of dictionaries have historically excluded encyclopedic content, although there are numerous outliers and the line separating lexical and encyclopedic information has never been clearly drawn.

For instance, the names of organizations with well-established metonymic applications are typically included in dictionaries. As a result, any headword list will typically include the Pentagon and Buckingham Palace because corpus data regularly contains sentences similar to these:

It appears that neither the Pentagon nor the Bush White House had prepared for such a situation.

In this instance, how can we be certain that Buckingham Palace has acted appropriately?

The same holds true for both actual and made-up locations with deep symbolic implications, such Shangri-la (a weekend spent in New York's homosexual Shangri-la) and Mecca (which has Sharm-el-Sheikh, a Mecca for sun worshippers and divers). Dictionary definitions usually cover just the extended usage in situations such as these. Similar to this, most dictionaries merely list names like Google and Facebook as verbs, without defining the actual nouns from which they originate.

For instance, if we choose to define nations (as well as languages and ethnicities, as is now customary), why not include cities as well, and where do we draw the line? If nations and cities are involved, then why not individuals — and if yes, which ones? The question of which named entities belong in a dictionary should be reexamined in light of the current situation and user expectations, even though the distinction between lexical and encyclopedic material is already becoming increasingly hazy in sites like Babelnet.

As Hanks has observed "The very word definition implies identifying boundaries" (Hanks 2013: 85), and this reflects a traditional view of word meaning which is now being challenged. The assumption that meanings are fixed entities, which "can be attributed to the word in isolation, rather than in context" (Hanks 2015: 87) and which can be described in terms of "necessary and sufficient conditions", is undermined by research in lexical semantics and prototype theory, backed up by the findings of corpus linguistics. This has given rise to new methods for defining—or more accurately, elucidating—word meanings. An emphasis on context and co-text is prevalent, and some of these experiments stretch back to the digital era.

On a lesser scale, the Macmillan Dictionary established a paradigm for using a second phrase to convey connotative (or pragmatic) information. In this instance, the denotative meaning of a term is explained by its conventional definition, while a second sentence provides information about the speaker's attitude or motivation for using the word.

FOR EXAMPLE:

bureaucrat someone who is employed to help run an office or government department. This word can suggest that you do not like people like this because you think they have too much power and care too much about rules and systems.

Example sentences are helpful to dictionary users. In addition to providing models for language production (particularly helpful in educational dictionaries), they also aid in the clarification of meanings and contextual choices (e.g., Atkins and Rundell 2008: 452-455). Long before dictionaries were available online, real examples in the form of whole phrases extracted from a corpus were replacing the previous paradigm of made-up, frequently abbreviated examples.

Since the 1990s, dictionaries available on optical disks (CD-ROM and DVD-ROM) have consistently included extra instances. These would normally be drawn from a corpus, but most of the time there was little to no screening (for appropriateness, quality, etc.), and—most importantly—examples of polysemous terms were not linked to particular senses.

"The dynamic nature of e-dictionaries enables lexicographers to move away from a static to a dynamic data display that includes the use of a multi-layered structure of dictionary articles" (Gouws 2014: 164).

It appears that several of the inventions discussed here have made logical steps in this regard. One trend is the growing adoption of general norms (rather than dictionary-specific ones) for information presentation and linking: tabs, hyperlinks, icons for expanding and collapsing a particular category of information, and so forth. The Web is becoming more standardized overall, and users are increasingly expected to be conversant with a certain "vocabulary" of search techniques.

So, it makes obvious sense for these to be used in dictionary sites, too, since the data on how people arrive at dictionary sites shows that — for many users — the destination is simply an outcome of search, rather than an instance of "looking it up in the dictionary". More generally, information scientists and new Web-oriented dictionary-user studies are valuable sources of information for dictionary builders. However, it's also been observed that a lot of these structural advances are used on out-of-date material. Even a resource as innovative as Babelnet relies mostly on Wiktionary for its dictionary information; frequently, the definitions of common words found there are derived from sources that date back more than a century.

As the scope of the dictionary expands and its structures develop to fully exploit the possibilities of digital media, the lexical data it delivers should also reflect the most up-to-date linguistic thinking about how humans create and understand meanings.

RESOURSES

1. Atkins, B.T.S. and M. Rundell. 2008. The Oxford Guide to Practical Lexicography. Oxford/New York: Oxford University Press.

2. Atwood, M. 2015. We are Double-plus Unfree. The Guardian, September 2015. http://www.theguardian.com/books/2015/ sep/18/margaret-atwood-we-are-double-plus-unfree.

3. Fatsis, S. 2015. The Definition of a Dictionary. Slate Magazine. January 2015. http://www.slate.com.

4. Gouws, R.H. 2014. Article Structures: Moving from Printed to e-Dictionaries. Lexikos 24: 155-177.

5. Hanks, P. 2010. Lexicography, Printing Technology, and the Spread of Renaissance Culture.

6. Dykstra, A. and T. Schoonheim (Eds.). 2010. Proceedings of the XIV EURALEX International Congress, Leeuwarden, 6–10 July, 2010: 988-1006. Leeuwarden/Ljouwert: Fryske Akademy.

7. Hanks, P. 2013. Lexical Analysis: Norms and Exploitations. Cambridge, Mass.: MIT Press.

8. Hanks, P. 2015. Cognitive Semantics and the Lexicon (Review Article). International Journal of Lexicography 28(1): 86-106.

9. Lew, R. 2014. User-generated Content (UGC) in English Online Dictionaries. OPAL:.

10. Lew, R. Forthcoming. Space Restrictions in Paper and Electronic Dictionaries and their Implications for the Design of Production Dictionaries., P. and B. Wójtowicz (Eds.). Forthcoming. Issues in Bański Modern Lexicography. München: Lincom Europa.

11. Rundell, M. 2006. More than One Way to Skin a Cat: Why Full-sentence Definitions have not been 1Universally Adopted. Corino E., C. Marello and C. Onesti (Eds.). 2006. Proceedings of the 12th

EURALEX International Congress, Torino, Italia, September 6–9, 2006: 323-337. Alessandria: Edizioni Dell'Orso.

12. Rundell, M. Forthcoming. Crowdsourcing, Wikis, and User-generated Content, and their Potential Value for Dictionaries. Hanks, P.W. and G.-M. de Schryver (Eds.). Forthcoming. International

13. Handbook of Modern Lexis and Lexicography. Berlin: Springer.

Schoonheim, T. and R. Tempelaars. 2010. Dutch Lexicography in Progress: The Algemeen Nederlands Woordenboek (ANW). Dykstra, A.and T. Schoonheim (Eds.). 2010. Proceedings of the XIV EURALEX International Congress, Leeuwarden, 6–10 July, 2010: 718-725.

14. Simpson, J. 2014. What Would Dr Murray Have Made of the OED Online Today 2.0, 2 (2): 15-36. Available at http://www.trojina.org/slovenscina2.0/arhiv/2014/2/Slo2.0_2014_2_015.

15. Stamper, K. 2015. This Wild and Barbarous Jargon, Reduced: Practical Lexicography in an Age of Digital Abundance. Video of talk given at eLex 2015: Electronic Lexicography in the 21st Century. https://elex.link/elex2015/videos/.

DICTIONARIES AND OTHER ONLINE RESOURCES:

1. Babelnet: http://babelnet.org

2. Macmillan Dictionary: http: //www.macmillandictionary.com.

3. Merriam-Webster online: http: //www.merriam-webster.com. Oxford Dictionaries Online http://www.oxforddictionaries.com.

4. Oxford Advanced Learner's Dictionary Online: www.oxfordlearnersdictionaries.com.

5. Pattern Dictionary of English Verbs (PDEV): http: //pdev.org.uk. Wiktionary: https://en.wiktionary.org/.

6. Wordnik: https://www.wordnik.com