Gilles-Maurice de Schryver

METALEXICOGRAPHY: AN EXISTENTIAL CRISIS

Abstract   While there was arguably a need for multi-authored, multi-volume, metalexicographic handbooks three decades ago – when the field of metalexicography was still 'young' – it is a bit puzzling to make sense of the current output-flurry in this field. Is it simply a matter of 'every publisher trying to fill its shelves'? or is there really a need in the scientific community for more and (continuously) updated reference works? And once available, are such works also consulted? Which parts? By whom? How often? For what purposes? In this paper we look at an ongoing, real-world metalexicographic handbook project to answer these questions.

Keywords   Metalexicography; major reference work; publishing model; download vs. citation patterns

1.   The booming field of major metalexicographic reference works

When HSK 5.1 to 5.3 was published by Walter de Gruyter (eds. Hausmann et al. 1989–1991), Robert Ilson, then the editor of IJL, could claim "I use this enormous work all the time" (1997, p. 348); but apart from him, who was actually (and who has been) consulting this encyclopaedia (which runs over 3,400 pages)? Soon enough, the need for a fourth volume was felt, to fill in gaps that had been missed and to take account of the computer revolution; HSK 5.4 (with another 1,600 pages) eventually appeared, after a publication delay of five years, two decades later (eds. Gouws et al. 2013).

Fast-forward to the 2010s, and the community is at it again: Continuum published *e-Lexicography* (eds. Fuertes-Olivera/Bergenholtz 2011; 350 pages), and OUP *Electronic Lexicography* (eds. Granger/Paquot 2012; 500 pages). Then came *The Bloomsbury Companion to Lexicography* (ed. Jackson 2013; 450 pages), *The Oxford Handbook of Lexicography* (ed. Durkin (2016); 700 pages), *The Routledge Handbook of Lexicography* (ed. Fuertes-Olivera 2018; 850 pages), and most recently *The Cambridge Companion to English Dictionaries* (ed. Ogilvie (2020); 400 pages). On the day this contribution is being finalised,[1] the second edition of *The Bloomsbury Handbook of Lexicography* (ed. Jackson 2022; 510 pages) has just been released: alongside significantly updated and thoroughly revised chapters, it now includes a further six new chapters. Expected later in the year, is the *Cambridge Handbook of the Dictionary* (eds. Finegan/Adams 2022).

2.   Case study: the *International Handbook of Modern Lexis and Lexicography* (IHMLL)

Crosscutting all these efforts is the *International Handbook of Modern Lexis and Lexicography* (IHMLL) (eds. Hanks/de Schryver 2014–2022), published by Springer as a so-called 'living reference work'. Initially planned to contain about 115 chapters (ranging from 10 to 80 pages each), chapters started appearing online in 2014 through 2017, after which the project stalled. Just 28 chapters had been finalised, or about a quarter (https://link.springer.com/

---

[1]   March 24, 2022.

referencework/10.1007/978-3-642-45369-4). There are quite a number of reasons for the standstill, but one of them, perhaps even the chief reason among them, was the result of an existential crisis: "What and who is all this really for?" The editors basically work for free, and all the authors contribute for free, while the publisher outsources the production to the cheapest corners of the world, necessitating endless revisions to multiple sets of proofs; yet once a chapter is finally 'ready' the publisher merely adds it to their ever-growing databases which contain literally hundreds of handbooks, totalling thousands of chapters, which they then bundle together with large numbers of digital books into 'eBook packages' for which academic libraries have to (and do!) pay eye-watering amounts, year after year.

2.1 Modern typesetting

There is clearly something very wrong with this modern business model; so wrong that it quite literally drove Patrick Hanks – otherwise no stranger to huge lexicographic projects – rather mad. The fact that such undertakings are often vanity projects, performed as an act of love for the field without proper remuneration is well known. But what is surely a new development is the extreme carelessness with which authors' texts are treated by typesetters. Never mind that many non-Latin letters (Greek, Cyrillic, even diacritics in languages such as Turkish or Vietnamese) come back garbled in the first set of proofs, an endless succession of 'project managers' at publishing houses is simply pushing buttons and handing down mindless instructions without anyone still actually checking anything for contents. Addenda 1 and 2 show two random examples: The first is an extract of page upon page in second (yes, second!) proofs filled with unreadable URLs; the second is an example of the mindboggling consequences of publishers employing people for the production process who clearly refuse to (or perhaps even cannot?) read. Proofs with utter gibberish are being presented with a straight face, and no amount of complaining seems to have any effect. One's project also moves from continent to continent, with new project managers being assigned every few months, so that the provocations never stop. It seems as if editors have to be grateful that big-name publishers are even willing to take on their projects, with the new normal now being that editors have to re-edit the work of the typesetters. Proofs are often so bad that editors 'hide' them from the authors, only sending through the second or third set of proofs for a final read.[2]

2.2 Publish as you go

Incomplete or not, this hasn't stopped the publisher to already include the IHMLL in its eBook packages, so that about a hundred libraries currently have access to the released quarter (https://worldcat.org/identities/lccn-n83-198535/).

Ironically, that released quarter led a life of its own over the past four years, as it has been available not only in the publisher's databases, but also in some of the authors' university repositories, as well as in academic social networks like ResearchGate (https://www. researchgate.net/) or Academia (https://www.academia.edu/). The four years that have elapsed since the stall thus provide us with natural access data.

---

[2] The problems with the way today's publishers behave, their new business models, and how they handle major reference works are not at all specific to metalexicography, alas.

2.3   Bibliometrics

In an earlier bibliometric study (to appear in 2022) of the lexicographic journals *Dictionaries*, IJL, *Lexikos* and *Lexicographica*, it was found that 40% of the articles published in those journals are never ever cited. Two out of every five journal articles in lexicography could just as well never have been written, as no one ever refers to them. A cynic could even say, given that no colleague ever feels the need to refer to that material, that their only purpose was to populate one's own CV. In order to do better, we should all write less – 40% less. The problem, of course, is to know which two out of every five journal articles not to write.[3]

Moving from what are typically one-off contributions in lexicographic journals to chapters in major metalexicographic reference works, the question becomes: If two out of every five articles in lexicography don't attract even a single citation, what is the situation for chapters in handbooks on lexicography? Such handbooks are more akin to proper dictionaries, so whether or not a certain chapter ends up being cited (akin to whether or not a certain lemma ends up being looked up), is probably even more hit and miss. The preliminary assumption is therefore that the number of chapters never cited in a major handbook will be even higher than 40%.

Surprisingly, the data – for which, see Addendum 3 – reveal otherwise. According to Google Scholar (https://scholar.google.com/) just 9 of the 28 chapters have not (yet) been cited, or thus 32%. It is important to note, however – and when focusing on citations only, one tends to miss this important point – that zero citations doesn't mean that those chapters are never looked at. Indeed, the metrics on the publisher's website indicate that *all* chapters have been downloaded at least about a hundred times so far, up to nearly six hundred times for the most frequently downloaded ones. Hence, chapters which may not have attracted any citations so far, such as those on Icelandic, Yiddish, or etymology, have all been downloaded around a hundred times. We may thus assume that also those contributions were not written in vain after all. The three most downloaded chapters are 'Dictionaries as aids for language learning' (574 x), 'Historical principles vs. synchronic approaches' (566 x), and 'The lexicography of Khmer' (468 x). The three most cited chapters are 'Dictionaries and their users' (30 x), 'Dictionaries and crowdsourcing, wikis and user-generated content' (19 x), and 'Bilingual lexicography: translation dictionaries' (19 x).

The correct way to look at the values, though, is to take the actual number of days each chapter has been online into account, which is why Addendum 3 also includes columns with the download and citation values expressed 'per year', meaning 'per 365 days'. Note that the rows of Addendum 3 are ordered, first, from highest to lowest 'citations/year', and second, for those without any citations, from highest to lowest 'downloads/year'.

2.4   Downloads vs. citations

Modern dictionaries are based on corpus data, and an interesting metalexicographical discussion of the past 15 years concerns the question as to whether or not there is a correlation between corpus frequency and look-up frequency. While initial results using Bantu data

---

[3]   While this figure may be shocking to many, it is actually pretty respectable for the humanities, where up to 82% of published articles may go uncited (see Remler 2014, citing evidence from Larivière et al. 2009). For more on the power-law behaviour in the distribution of citations, see for instance Brzezinski (2015).

suggested that there is no useful relationship (de Schryver et al. 2006), this was challenged using a better methodology with Indo-European data (Koplenig/Meyer/Müller-Spitzer 2014; Trap-Jensen/Lorentzen/Sørensen 2014; Müller-Spitzer/Wolfer/Koplenig 2015), up-on which two of the teams joined hands, to indeed conclude that there is a positive correlation after all, one which they now even claim is universal (de Schryver/Wolfer/Lew 2019).

A parallel question may also be asked of the IHMLL, namely: *To what extent do download numbers predict citation patterns in metalexicography?* At face value, one would assume that more downloads lead to more citations, and that fewer downloads will inevitably result in fewer citations. Researchers have for instance shown this to be the case in the field of analytical chemistry (Jahandideh/Abdolmaleki/Barzegari Asadabadi2007), but our (admittedly very limited) data for the field of metalexicography do not seem to corroborate that. Figure 1 brings the two values together for the first 28 chapters of the IHMLL: on the *x*-axis the number of downloads/year, on the *y*-axis the number of citations/year. The top three downloaded chapters are all found in the bottom-right of the graph, where they barely attracted citations. Conversely, it is those chapters with medium numbers of downloads/year, i. e. those in the top half of the graph, which resulted in the most citations/year. The 9 chapters with no citations of course pull the trendline down, revealing one additional outlier, viz. the chapter on 'Natural Semantic Metalanguage and lexicography'.

Overall, and over the past few years, each chapter was downloaded an average of 45 times a year, and cited an average of 0.74 times a year. And this while the project has been basically dormant, with no advertisement of it whatsoever, nor even a word about its very existence.[4]

---

4   The sum of the downloads for the 28 chapters in Addendum 3 is 7,487, while the Springer website claims "9.2k" for this metric. Following an enquiry, the Springer 'project coordinator' for the IHMLL at the time confirmed that "due to a temporary bug" the overall metric shown online is about 2,000 downloads too high (personal communication, e-mail 16/02/2022).
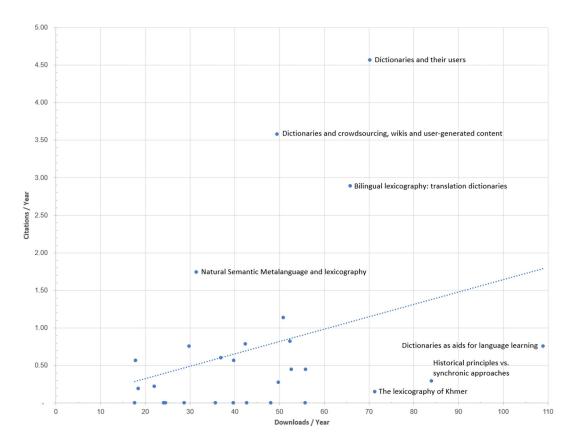
Fig. 1:   Downloads/year vs. citations/year for the IHMLL

The positioning of the dots in the graph of Figure 1 is intriguing, and begs for a comparison with more general trends in metalexicography. For earlier work in this field, see de Schryver (2009a, 2009b, 2012a, 2012b, 2019) and Lew/de Schryver (2014). The quickest 'single-shot, one-stop' place to assess Figure 1 is probably to allow Google Scholar (GS) to do the heavy lifting. Bringing the world's lexicographers with a GS profile (and the label 'lexicography') together, as done here https://scholar.google.com/citations?view_op=search_authors&hl=en &mauthors=label:lexicography, immediately reveals that most citations in metalexicography are collected by those colleagues involved in the interface between language teaching and lexicography (cf. Batia Laufer and Sylviane Granger in positions 1 and 3). The users of the IHMLL follow this pattern in terms of top download (cf. 'Dictionaries as aids for language learning') but, as seen above, do not follow this up with citations. Similarly, 'Historical principles vs. synchronic approaches' is an old favourite in traditional metalexicography (cf. Patrick Hanks in position 2), again not followed up in terms of citations. Conversely, the new metalexicographic topics which focus on digital users, crowdsourcing, etc. are only found on pages two and three (each page lists 10 colleagues) at GS (cf. e.g. Michael Rundell and Robert Lew). What this suggests is that the user of the IHMLL consults this handbook to move the modern aspects of our field forward, engaging with and citing the material, while the older topics, while consulted (proxy: 'downloaded') more often, are mainly browsed out of interest.

3.   Outlook

In conclusion we can thus state that handbooks such as the IHMLL do have an audience and a purpose after all. Work on the IHMLL has therefore restarted in earnest: Over the past half year a further 22 chapters were put into production, which will bring the total to 50. By the time of the Euralex 2022 congress, we expect to be closer to the end. Here's hoping.

References

Brzezinski, M. (2015): Power laws in citation distributions: evidence from scopus. In: Scientometrics 103 (1), pp. 213–228.

de Schryver, G.-M. (2009a): Bibliometrics in lexicography. In: International Journal of Lexicography 22 (4), pp. 423–465.

de Schryver, G.-M. (2009b): Lexikos at eighteen: an analysis. In: Lexikos 19, pp. 372–403.

de Schryver, G.-M. (2012a): Lexicography in the crystal ball: facts, trends and outlook. In: Fjeld, Ruth V./Torjusen, Julie M. (eds.): Proceedings of the 15th EURALEX International Congress, 7–11 August, 2012, Oslo. Oslo, pp. 93–163.

de Schryver, G.-M. (2012b): Trends in twenty-five years of academic lexicography. In: International Journal of Lexicography 25 (4), pp. 464–506 + 42 pages of supplementary material online.

de Schryver, G.-M. (2019): Past, present and future in Asian lexicography. Paper presented at the 13th International Conference of the Asian Association for Lexicography, ASIALEX 2019, Istanbul, Turkey, 19–21 June 2019.

de Schryver, G.-M./Wolfer, S./Lew, R. (2019): The relationship between dictionary look-up frequency and corpus frequency revisited: a log-file analysis of a decade of user interaction with a Swahili-English dictionary. In: GEMA Online® Journal of Language Studies 19 (4), pp. 1–27 + supplementary material online.

de Schryver, G.-M./Joffe, D./Joffe, P./Hillewaert, S. (2006): Do dictionary users really look up frequent words? – On the overestimation of the value of corpus-based lexicography. In: Lexikos 16, pp. 67–83.

Durkin, P. (ed.) (2016): The Oxford handbook of lexicography. New York.

Finegan, E./Adams, M. (eds.) (2022): The Cambridge handbook of the dictionary. Cambridge.

Fuertes-Olivera, P. A. (ed.) (2018): The Routledge handbook of lexicography. London.

Fuertes-Olivera, P. A./Bergenholtz, H. (eds.) (2011): e-Lexicography: the internet, digital initiatives and lexicography. London.

Gouws, R. H./Heid, U./Schweickard, W./Wiegand, H. E. (eds.) (2013): Dictionaries. An international encyclopedia of lexicography. Supplementary volume: Recent developments with focus on electronic and computational lexicography. (= Handbooks of Linguistics and Communication Science (HSK) 5.4). Berlin.

Granger, S./Paquot, M. (eds.) (2012): Electronic lexicography. Oxford.

Hanks, P./de Schryver, G.-M. (eds.) (2014–2022): International handbook of modern lexis and lexicography. Berlin.

Hausmann, F. J./Reichmann, O./Wiegand, H. E./Zgusta, L. (eds.) (1989–1991): Wörterbücher/ Dictionaries/Dictionnaires. Ein internationales Handbuch zur Lexikographie/An international encyclopedia of lexicography/Encyclopédie internationale de lexicographie. (= Handbücher zur Sprach- und Kommunikationswissenschaft (HSK) 5.1–5.3). Berlin.

Ilson, R. F. (1997): Review: Wörterbücher/Dictionaries/Dictionnaires: An international encyclopedia of lexicography. Edited by Franz Josef Hausmann, Oskar Reichmann, Herbert Ernst Wiegand, Ladislav Zgusta. Berlin/New York. Vol. 1, 1989; Vol. 2, 1990; Vol. 3, 1991. pp. LII+3355. In: International Journal of Lexicography 10 (4), pp. 348–357.

Jackson, H. (ed.) (2013): The Bloomsbury companion to lexicography. London.

Jackson, H. (ed.) (2022): The Bloomsbury handbook of lexicography (= Bloomsbury Handbooks). London.

Jahandideh, S./Abdolmaleki, P./Barzegari Asadabadi, E. (2007): Prediction of future citations of a research paper from number of its internet downloads. In: Medical Hypotheses 69 (2), pp. 458–459.

Koplenig, A./Meyer, P./Müller-Spitzer, C.n (2014): Dictionary users do look up frequent words. A log file analysis. In: Müller-Spitzer, C. (ed.). Using Online Dictionaries (= Lexicographica. Series Maior 145). Berlin, pp. 229–249.

Larivière, V./Gingras, Y./Archambault, É. (2009): The decline in the concentration of citations, 1900–2007. In: Journal of the American Society for Information Science and Technology 60 (4), pp. 858–862.

Lew, R./de Schryver, G.-M. (2014): Dictionary users in the digital revolution. In: International Journal of Lexicography 27 (4), pp. 341–359.

Müller-Spitzer, C./Wolfer, S./Koplenig, A. (2015): Observing online dictionary users: studies using Wiktionary log files. In: International Journal of Lexicography 28 (1), pp. 1–26.

Ogilvie, S. (ed.) (2020): The Cambridge companion to English dictionaries. Cambridge.

Remler, D. (2014): Are 90% of academic papers really never cited? Searching citations about academic citations reveals the good, the bad and the ugly.
https://dahliaremler.com/2014/04/09/are-90-of-academic-papers-really-never-cited-searching-citations-about-academic-citations-reveals-the-good-the-bad-and-the-ugly/.

Trap-Jensen, L./Lorentzen, H./Sørensen, N. H. (2014): An odd couple – corpus frequency and look-up frequency: what relationship? In: Slovenščina 2.0 2 (2), pp. 94–113.

Contact information

Gilles-Maurice de Schryver
BantUGent – UGent Centre for Bantu Studies, Ghent University
&
Department of African Languages, University of Pretoria
gillesmaurice.deschryver@UGent.be

XX EURALEX

Addendum 1: Example of unreadable second proofs

119 the Greek language of the seventeenth to nineteenth centuries. There are some
120 valuable resources for that period such as Johannes Meursius' *Glossarium*
121 *Graeco-Barbarum* (1614) (more at https://en.wikipedia.org/wiki/Johannes_
122 Meursius), Alessio da Somavera's *Tesoro della lingua greca-volgare ed italiana*
123 (1709) (which can be accessed at http://webcache.googleusercontent.com/search?q=
124 cache:oglfkpQdkDsJ:anemi.lib.uoc.gr/metadata/2/4/b/metadata-01-0001129.tkl+&
125 cd=1&hl=el&ct=clnk&gl=gr), and Skarlatos Vyzantios' Dictionary (1835) (which
126 can be accessed at http://anemi.lib.uoc.gr/search/?dtab=m&search_type=simple
127 &search_help=&display_mode=overview&wf_step=init&show_hidden=0&number=
128 10&keep_number=10&cclterm1=&cclterm2=&cclterm3=&cclterm4=&cclterm5=&
129 cclterm6=&cclterm7=&cclterm8=&cclfield1=&cclfield2=&cclfield3=&cclfield4=&
130 cclfield5=&cclfield6=&cclfield7=&cclfield8=&cclop1=&cclop2=&cclop3=&cclop4=
131 &cclop5=&cclop6=&cclop7=&isp=&display_help=0&offset=11&search_coll%5b
132 metadata%5d=1&&stored_cclquery=creator%3D%28%CE%92%CF%85%CE%B6%
133 CE%AC%CE%BD%CF%84%CE%B9%CE%BF%CF%82%2C+%CE%A3%CE%
134 BA%CE%B1%CF%81%CE%BB%CE%AC%CF%84%CE%BF%CF%82+%CE%
135 94.%2C%29&skin=&rss=0&show_form=&export_method=none&ioffset=1&
136 display_mode=detail&ioffset=1&offset=13&number=1&keep_number=10&old_
137 offset=11&search_help=detail), et al. Stefanos Koumanoudis attempts to gather in
138 his "Collection" (1900) (which can be accessed at http://anemi.lib.uoc.gr/search/?
139 dtab=m&search_type=simple&search_help=&display_mode=overview&wf_step=
140 init&show_hidden=0&number=10&keep_number=10&cclterm1=&cclterm2=&
141 cclterm3=&cclterm4=&cclterm5=&cclterm6=&cclterm7=&cclterm8=&cclfield1=&
142 cclfield2=&cclfield3=&cclfield4=&cclfield5=&cclfield6=&cclfield7=&cclfield8=&
143 cclop1=&cclop2=&cclop3=&cclop4=&cclop5=&cclop6=&cclop7=&isp=&display_
144 help=0&offset=11&search_coll[metadata]=1&&stored_cclquery=creator=(Κουμανο

Addendum 2: Example of typesetters sleeping at the wheel

Author input:

211 *Perkamusan Indonesia* (1976) lists no less than 16 post-independence dictionaries
212 intended for an Indonesian readership, and many more have appeared since. Among
213 ==the== ==most== ==exhaustive== ==is== ==a== ==756-page== Javanese-Indonesian dictionary by
214 Prawiroatmodjo (1957).

> **user**
>
> layout?

Typesetter action:

Since Indonesia's independence, Javanese remained one of the best documented languages, second only to Malay/Indonesian. The aforementioned *Bibliografi Perkamusan Indonesia* (1976) lists no less than 16 post-independence dictionaries intended for an Indonesian readership, and many more have appeared since. Among layout? Javanese-Indonesian dictionary by Prawiroatmodjo (1957).

Addendum 3: Download and citation data for the IHMLL so far (28 chapters, up to 24/03/2022)

Chapter title	Author(s)	Online since	Days online	Down-loads up to 24/03/2022	Down-loads / year	Citations up to 24/03/2022	Citati-ons / year
Dictionaries and their users	Robert Lew	03/09/2015	2,400	462	70.26	30	4.56
Dictionaries and crowdsourcing, wikis and user-generated content	Michael Rundell	07/12/2016	1,939	263	49.51	19	3.58
Bilingual lexicography: translation dictionaries	Arleta Adamska-Sałaciak	03/09/2015	2,400	433	65.85	19	2.89
Natural Semantic Metalanguage and lexicography	Cliff Goddard	30/08/2017	1,673	144	31.42	8	1.75
Construction grammar and lexicography	Bill Croft/Logan Sutton	21/12/2016	1,925	268	50.82	6	1.14
Term banks	Thierry Fontenelle/Dieter Rummel	06/12/2014	2,671	383	52.34	6	0.82
Figurative language and lexicography	Alice Deignan	17/11/2015	2,325	270	42.39	5	0.78
Dictionaries as aids for language learning	Alex Boulton/Sylvie De Cock	22/12/2016	1,924	574	108.89	4	0.76
The lexicography of Sephardic Judaism	David M. Bunis	23/12/2016	1,923	157	29.80	4	0.76
The lexicography of German	Annette Klosa	06/04/2017	1,819	184	36.92	3	0.60
Sign language lexicography	Rachel McKee/Mireille Vale	22/12/2016	1,924	210	39.84	3	0.57
The lexicography of Portuguese	Ana Frankenberg-Garcia	20/12/2016	1,926	94	17.81	3	0.57
The lexicography of Indonesian/Malay	Deny Arnos Kwary/ Nor Hashimah Jalaluddin	14/07/2015	2,451	375	55.84	3	0.45
The lexicography of indigenous languages in Australia and the Pacific	Nick Thieberger	01/08/2015	2,433	351	52.66	3	0.45

Chapter title	Author(s)	Online since	Days online	Downloads up to 24/03/2022	Downloads / year	Citations up to 24/03/2022	Citations / year
Historical principles vs. synchronic approaches	Judy Pearsall	06/07/2015	2,459	566	84.01	2	0.30
The lexicography of indigenous languages in South America	Wolf Dietrich	06/12/2014	2,671	364	49.74	2	0.27
Norms and exploitations in lexicography	Sara Može	13/09/2017	1,659	100	22.00	1	0.22
The lexicography of Scots	Susan Rennie	11/01/2017	1,904	96	18.40	1	0.19
The lexicography of Khmer	Robert K. Headley	08/09/2015	2,395	468	71.32	1	0.15
The lexicography of minority languages in Southeast Asia	David Bradley	03/07/2015	2,462	376	55.74	0	0
The lexicography of Persian (Farsi, Tajiki, and Dari)	Corey Andrew Miller	25/11/2017	1,586	209	48.10	0	0
The lexicography of Hebrew	Tsvi Sadan	10/01/2017	1,905	223	42.73	0	0
The lexicon of the male sex worker	Welby Ings	13/11/2015	2,329	253	39.65	0	0
The lexicography of Norwegian	Oddrun Grønvik	21/11/2016	1,955	191	35.66	0	0
The lexicography of Esperanto	Federico Gobbo	03/12/2016	1,943	153	28.74	0	0
Etymology in dictionaries	Anatoly Liberman	06/09/2017	1,666	112	24.54	0	0
The lexicography of Yiddish	Wolf Moskovich	21/06/2017	1,743	115	24.08	0	0
The lexicography of Icelandic	Þórdís Úlfarsdóttir/Kristín Bjarnadóttir	20/12/2016	1,926	93	17.62	0	0
	AVERAGES		2,083	267	45.24	4	0.74