STATE-OF-THE-ART OF MONOLINGUAL LEXICOGRAPHY FOR SWEDEN

Emma SKÖLDBERG, Louise HOLMER, Elena VOLODINA, Ildikó PILÁN

Department of Swedish, University of Gothenburg

Sköldberg, E., Holmer, L., Volodina, E, Pilán, I. (2019). State-of-the-art on monolingual lexicography for Sweden. Slovenščina 2.0, 7 (1): 13–24.

DOI: http://dx.doi.org/10.4312/slo2.0.2019.1.13-24.

The minireview describes the state-of-the-art of Swedish monolingual lexicography. The main actors in the field, both commercial and non-commercial, are mentioned alongside with the description of lexicographic products that have been offered by them to the lexicon users. The minireview makes it clear that there is an obvious tendency among the Swedish dictionary users to abandon paper-based dictionaries and switch over to online portals and apps, which influences the practices adopted by commercial publishing houses, such as Norstedts, Bonniers, Natur & Kultur. Among the leading non-commercial players, the Swedish Academy, the Swedish Language Bank, Institute for Language and Folklore are named. Swedish monolingual lexicography offers, however, dictionaries produced not only by experts but also by non-experts (i.e. using the efforts of the crowd).

Keywords: Swedish monolingual lexicography, Swedish lexicographic actors, Swedish online lexicographic portals, Swedish UGC dictionaries

1 THE STATUS OF SWEDISH LEXICOGRAPHY AND DICTIONARIES IN GENERAL

The status of lexicography in Sweden does probably not differ from the situation in many other countries. The dictionary users are interested in solving a problem – learn about the meaning or usage of a word, translating a word, etc. – and they expect to be supported directly, via their digital device and at no cost. From a lexicographical point of view, it is preferable if users are supported by dictionaries. But probably a great part of the users do not separate between different types of resources, such as professionally compiled dictionaries, Google translate, built-in spelling tools, etc. These circumstances, of course, do influence the modern dictionary market to a great extent.

The former giants on the Swedish commercial dictionary market, namely the traditional publishing companies, such as Norstedts,¹ Natur & Kultur,² and Bonniers,³ proved to be in a period of prosperity approximately between 1990 and 2005 (Sköldberg and Mattsson 2016), but now the commercial market for print dictionaries is strongly reduced. Nowadays, the publishing companies focus on accumulating their resources (i.e. digitized paper-based dictionaries) in different web portals intended for education (like ord.se, ne.se, WordFinder.com, etc.). Given the competitive situation, it is not easy for outsiders to get information about the number of subscribers and active users of these commercial sites.

At the same time, the Swedish Academy⁴ plays a unique role in the Swedish dictionary context, being a private, non-profit foundation since the year of 1786. The Swedish Academy finances three major dictionaries: *The Swedish Academy Dictionary* (SAOB), *The Swedish Academy Glossary* (SAOL) and the *Contemporary Dictionary of the Swedish Academy* (SO); all three included in the dictionary portal svenska.se (see below).

Another actor on the Swedish "dictionary market" is the governmental Institute for Language and Folklore,⁵ which produces dialect dictionaries based

- 1 http://www.norstedts.se/
- 2 https://www.nok.se/
- 3 https://www.bonnier.com/en/
- 4 https://www.svenskaakademien.se/en
- 5 http://www.sprakochfolkminnen.se/

on extensive collections of dialect transcriptions and recordings (Språk- och folkminnen, 2017). This Institute also offers official language counseling, and it provides the monolingual Swedish learner's dictionary *Lexin*, targeting relatively advanced beginner learners of Swedish. The monolingual version of Lexin has also been the base for about 20 bilingual dictionaries (see below).

The Swedish Language Bank, established already in 1975, and recently granted a status of national infrastructure for providing and maintaining openly available lexical and textual resources, provides several freely available lexical resources. Even though many of the resources are primarily intended for use in language technology applications, they are still very much lexicographical entities. Unique lexicons like SALDO (Borin et al., 2013), Constructicon (Lyngfelt et al. 2018), Swedish FrameNet (Ahlberg et al., 2014), Swesaurus (Borin and Forsberg 2014) and the SIMPLE lexicon (Kokkinakis et al., 2000), often created as a part of research projects, are made freely available through Språkbanken. The lexical search engine Karp (Ahlberg et al., 2016, Borin et al. 2012a) makes it possible to perform searches in several lexicons at the same time. Besides access to lexicons, Språkbanken collects, maintains and provides access to extensive collections of Swedish texts, i.e. corpora. To Swedish lexicographers, especially the ones working at the University of Gothenburg, the corpora in Språkbanken have played a crucial role (Malmgren and Sköldberg, 2013). The large-scale and diverse resources included in Korp (Borin et al., 2012b) can not only support linguistically interested users, but also form the basis of significant theoretical contributions to e.g. lexical semantics and lexical typology. Thus, the infrastructure to access corpora and different kinds of lexical resources in Sweden is in an advanced stage. However, it remains somewhat unclear to what extent Språkbanken's research and resources are known among the general public.

2 GENERAL MONOLINGUAL DICTIONARIES BY PROFESSIONALS

As mentioned above, The Swedish Academy plays an important role on the Swedish lexicographical scene. Its major commitment is *Ordbok över sven-ska språket utgiven av Svenska Akademien* (i.e. *The Swedish Academy Dictionary*; SAOB), a comprehensive historical dictionary describing the Swedish language from 1521 and to present time. The first volume of the dictionary

was published in 1898 and the latest, number 37, including words beginning with the letter V, was published in 2017. SAOB now holds more than 500,000 headwords. The SAOB is edited in Lund by a staff of about 20 employees. In a couple of years, the editorial board will have reached the last letter in the alphabet (Ö). The revision and update of previous volumes will then be intensified (Rosqvist, 2014; SAOB, 2018).

The most widely recognized dictionary published by the Academy is probably the contemporary Svenska Akademiens ordlista (i.e. The Swedish Academy Glossary; SAOL). SAOL was first published in 1874. A fully revised and updated edition, number 14, was published in 2015 (Malmgren, 2014). SAOL is the unofficial norm for orthography and inflection of contemporary Swedish words and comprises about 126,000 headwords. Even though the semantic description plays a minor role in the glossary, user comments and a study have shown that it is relatively often used to support tasks related to reception (Holmer et al., 2015b). In other words, the dictionary is influential in many ways among the users, but it does not have the same official status in Sweden as the Danish Retskrivningsordbogen (RO) in Denmark. The content of all 14 editions of SAOL (appr. 200,000 lemmas) have been digitized. Headwords (and facsimile) from all editions are free to access via the website SAOLhist.se, which offers excellent opportunities to study the development of the Swedish vocabulary over the past 140 years (Holmer, 2016; Holmer et al., 2016; Holmer and von Martens, 2016).

Finally, Svensk Ordbok utgiven av Svenska Akademien (i.e. The Contemporary Dictionary of the Swedish Academy; SO) was first published in 2009, and is now undergoing revision, aiming at being e-published in its second edition in 2020. SO holds about 65,000 headwords along with definitions, etymology, morphological and syntactical examples, style guides and more. SO is the newest of the Academy products and so far the least known among (potential) users (see further Malmgren, 2009; Holmer et al., 2015a). The editorial staff of the two contemporary dictionaries is employed by the University of Gothenburg.

The dictionaries published by the Swedish Academy are available in print format (all three of them) but also as apps for iOS and Android (SAOL and SO) and on a coordinated web portal (svenska.se; see below).

The SAOL app has been free of charge since its first launch in 2011. The SO app used to be a paid app (appr. 5 EUROS) when it was launched in 2015, but since Sept. 2017 it is also free of charge. The SAOL app in particular has had many downloads, about 800,000 altogether.

The owners of a whole set of print SAOB volumes are few in number, but the dictionary has also been freely available online since 1997. A new digital version, also for free, is now to be found at saob.se (Bäckerud, 2016). The interface of the website is user-friendly and should satisfy a large number of users. However, the older version offers more advanced search options which still makes it very relevant to specialists, such as language historians etc. (Petzell, 2017). The proportion of smartphone users of saob.se is quite large, as a result of the editors' decision to focus on a responsive web version.

In September 2017, the Swedish Academy launched its dictionary website, svenska.se. The website allows users to search for headwords in the three Academy dictionary products at the same time. The dictionaries complement each other as they i) cover different periods of time, ii) focus on different aspects of the Swedish vocabulary, and iii) differ in their descriptive/normative aims, respectively. However, only the information from the print editions is shown on the website, which means that relatively new words may not be described. Obviously, it would be practical if the dictionaries were to be updated on a regular basis. So far, there is little user statistics, but the usage is constantly increasing, and the awareness of and knowledge about the site svenska.se among the general public is steadily growing.

3 MONOLINGUAL DICTIONARIES BASED ON CROWDSOURCING

Of course, there are also Swedish dictionaries based on user-generated content (UGC), created with the help from users. International researchers often study Wiktionary (see e.g. Fuertes-Olivera, 2009; Meyer and Gurevych, 2012; Wolfer and Müller-Spitzer, 2016), but the Swedish version of Wiktionary does not seem to play a particularly central role among Swedish users, and it has so far, to our knowledge, not been the object of any scientific studies. However, content from Swedish Wiktionary has been used in language technology contexts (e.g. Borin and Forsberg, 2014). There are also other online dictionaries with UGC, for example Folkets lexikon (Kann 2010) and Folkmun.se (Sköldberg and

Wenner, 2017). The latter is a freely available dictionary that includes slang words and dialectal words, thus supplementing the more comprehensive dictionaries that describe style-neutral words used in major parts of the country.

Finally, in connection to previous studies of UGC dictionaries, the question is raised about how many *different* users that actually contribute to these dictionaries (see e.g. Wolfer and Müller-Spitzer, 2016). The basic idea is that "everyone can contribute" but the question is whether they do. Sköldberg and Wenner (2018) show that Folkmun.se in total includes about 5,500 entries (Dec. 2016) compiled by about 1,700 unique contributors, which may be considered a relatively satisfying diversity.

According to the Alexa ranking, one of the most frequented Swedish dictionary sites is synonymer.se. The site has been available online for a longer period of time than most of the Academy products, and it generally reaches a high page ranking when users consult a search engine for practically any word.

Synonymer.se combines lexicographically produced content with UGC. The site originally builds upon a print synonym dictionary (published by Bonniers, 2008). Online users are encouraged to add synonyms and new headwords. Some content origins from another UGC synonym dictionary (cf. Kann and Rosell, 2005). The site owners have also accessed the rights to the monolingual Swedish print dictionary *Bonniers svenska ordbok* (2010), adding its content to the site (see Holmer and Sköldberg, 2016).

4 SPECIAL MONOLINGUAL DICTIONARIES FOR SCHOOLS, FOREIGN LEARNERS ETC.

The specific national curricula for schools have few guidelines concerning the use of dictionaries for native speakers. Several Swedish schools have multi-user licenses on web portals including reference books and dictionaries (like ne.se, which for example includes the comprehensive *Nationalencyklopedins ordbok*, NEO, from 1995–96). Access to digital devices and the Internet is very high in Sweden in general. Some schools may of course (still) have class sets of a particular printed dictionary in their classrooms – hypothetically, as a way to train children in using paper-based versions of dictionaries.

One dictionary that has to be mentioned in this context is *Lexin* (also mentioned above). It is a learner's dictionary of Swedish, published for the first time in 1984 and available online since the late 1990's. The Swedish dictionary of *Lexin* holds about 30.000 headwords. The fourth and latest edition is only published online. The dictionary has had many users over the years, and although many of them belong to *Lexin*'s target group, many native speakers use *Lexin* as well. The use of *Lexin* online is investigated by Ann-Kristin Hult in her doctoral thesis (2016; see also Hult, 2014).

Lexin has in a way been complemented by another learner's dictionary from Natur & Kultur, Natur and Kulturs Svenska ordbok (2001), followed by a revised edition Natur and Kulturs Stora Svenska Ordbok (2006). These are published by a commercial company and are only available in print, whereas Lexin is a governmental project and product, and is freely available online. Several private initiatives have also developed smart phone apps for Android based on Lexin.

5 FINAL REMARKS

The current lexicographic situation in Sweden is somewhat ambiguous. On the one hand, the dictionary publishing houses have economic problems as fewer people invest in paper-based dictionaries nowadays compared to 50–20 years ago. The labour market for professional lexicographers is now very limited. Relatively few users are willing to pay for dictionary apps and licenses for Internet dictionaries. On the other hand, there are also positive exceptions on the Swedish lexicographic scene thanks to some non-commercial actors, especially the Swedish Academy and the governmental Institute for Language and Folklore. Furthermore, the Swedish Language Bank (University of Gothenburg) is a nationally and internationally acknowledged research unit in the field of language technology, with one of its special efforts aimed at supporting (as well as initiating and leading) lexicographic projects.

As the Swedish Academy finances monolingual works, the situation for monolingual dictionaries developed by professional lexicographers is better than for bilingual dictionaries, which traditionally have been compiled by publishing houses (cf. *Lexin* above). One of few exceptions is the site ord.se, which provides monthly subscribers with several bilingual dictionaries. However, some

more institutions involved in developing lexicographical products would have been preferable, to provide a wider and more varied description of the Swedish language.

REFERENCES

Dictionaries, databases and digital resources:

Bonniers svenska ordbok (2010). P. A. Sjögren, I. Györki & S. Malmström (Eds.). 10th ed. Stockholm: Bonniers.

Folkmun.se = David Eriksson (2018). Retrieved from https://www.folkmun.se/

Lexins svenska lexikon online. 4th ed. Retrieved from http://lexin.nada.kth.se/lexin/

Natur och Kulturs Stora Svenska Ordbok (2006). P.-E.Köhler & U. Messelius (Eds.). Stockholm: Natur & Kultur.

Natur och Kulturs Svenska Ordbok (2001). P.-E.Köhler & U. Messelius (Eds.). Stockholm: Natur & Kultur.

NEO = *Nationalencyklopedins ordbok* (1995–96). Höganäs: Bra Böcker.

Ne.se = *Nationalencyklopedin* (2018). Retrieved from https://www.ne.se/

Ord.se = Nationalencyklopedin (2018). Retrieved from https://ne.ord.se/

RO = *Retskrivningsordbogen* (2012). Jervelund, A. Ågerup, J. Schack, J. Norby Jensen and M. Heidemann Andersen (eds.). 4 ed. Copenhagen: Dansk Sprognaevn and Alinea A/S.

SAOB = *Ordbok över svenska språket utgiven av Svenska Akademien* (1898–). Lund: Gleerups. Retrieved from www.saob.se

SAOL 14 = *Svenska Akademiens ordlista över svenska språket* (2015). 14th ed. Stockholm: Norstedts i distribution.

SAOLhist 2018. Retrieved from http://spraakdata.gu.se/saolhist/

SO = Svensk ordbok utgiven av Svenska Akademien (2009). Stockholm: Norstedts.

Svenska.se = *Svenska Akademien* (2018). Retrieved from https://svenska.se/

Synonymer.se = *Sinovum Media* (2018). Retrieved from https://www.synonymer.se/

Wiktionary (2018). Retrieved from http://sv.wiktionary.org/wiki/Wiktionary:Huvudsida

WordFinder.com (2018). Retrieved from https://www.wordfinder.com

Other:

- Ahlberg, M., Borin, L., Forsberg M., Olsson, O., Schumacher, A., & Uppström J. (2016). Karp: Språkbanken's Open Lexical Infrastructure. In *Proceedings of Globalex 2016*.
- Ahlberg, M., Borin, L., Dannélls, D., Forsberg, M., Toporowska Gronostaj, M., Friberg Heppin K., Johansson, R., Kokkinakis, D., Olsson, L. J., & Uppström, J. (2014). Swedish Framenet++. The beginning of the end and the end of the beginning. In *Proceedings of the Fifth Swedish Language Technology Conference, Uppsala*, 13-14 November 2014.
- Borin, L., & Forsberg, M. (2014). Swesaurus; or, The Frankenstein approach to Wordnet construction. In *Proceedings of the Seventh Global WordNet Conference (GWC 2014)* (pp. 215–223).
- Borin, L., Forsberg, M., & Lönngren, L. (2013). SALDO: a touch of yin to WordNet's yang. *Language resources and evaluation*, 47(4), 1191–1211.
- Borin, L., Forsberg, M., Olsson, L. J., & Uppström, J. (2012a). The open lexical infrastructure of Språkbanken. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)* (pp. 3598–3602). Istanbul: ELRA.
- Borin, L., Forsberg, M., & Roxendal, J. (2012b). Korp the corpus infrastructure of Språkbanken. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)* (pp. 474–478). Istanbul: ELRA.
- Bäckerud, E. (2016). Indexeringen av SAOB. In A. Gudiksen & H. Hovmark (Eds.), *Nordiske studier i leksikografi 13* (pp. 101–109). Copenhagen.
- Fuertes-Olivera, P. A. (2009). The Function theory of lexicography and electronic dictionaries: WIKTIONARY as a prototype of collective free multiple-language Internet dictionary. In H. Bergenholtz, S. Nielsen & S. Tarp (Eds.), *Lexicography at a crossroads: dictionaries and encyclopedias today, lexicographical tools tomorrow* (pp. 99–134). Bern: Peter Lang.
- Holmer, L. (2016). *Grammatik i SAOL. En undersökning av grammatisk information i Svenska Akademiens ordlista över 130 år*. (Miss 65.) Göteborg.
- Holmer, L., Malmgren, S.-G., & von Martens, M. (2016). SAOLhist.se för allmänt och vetenskapligt bruk. In A. Gudiksen & H. Hovmark (Eds.), *Nordiske studier i leksikografi 13* (pp. 349–358). Copenhagen.

- Holmer, L., & von Martens, M. (2016). The SWEGOT Lexical Resources Framework. In T. Margalitadze & G. Meladze (Eds.), *Proceedings of the 17th EURALEX International Congress* (pp. 915–918). Tblisi: Ivane Javakhishvili Tbilisi University Press.
- Holmer, L., von Martens, M., & Sköldberg, E. (2015a). Making a dictionary app from a lexical database: the case of the *Contemporary Dictionary of the Swedish Academy*. In I. Kosem, M. Jakubíček, J. Kallas & S. Krek (Eds.), *Electronic lexicography in the 21st century: linking lexical data in the digital age. Proceedings of the eLex 2015 conference* (pp. 32–50). Ljubljana/Brighton: Trojina, Institute for Applied Slovene Studies/Lexical Computing Ltd.
- Holmer, L., Hult, A. K., & Sköldberg, E. (2015b). Spell-checking on the fly? On the use of a Swedish dictionary app. In I. Kosem, M. Jakubíček, J. Kallas & S. Krek (Eds.), *Electronic lexicography in the 21st century: linking lexical data in the digital age. Proceedings of the eLex 2015 conference* (pp. 356–371). Ljubljana/Brighton: Trojina, Institute for Applied Slovene Studies/Lexical Computing Ltd.
- Holmer, L., & Sköldberg, E. (2016). Synonymer.se i fokus om användningen av en svensk ordbokssajt. In A. W. Gustafsson, L. Holm, K. Lundin, H. Rahm & M. Tronnier (Eds.), *Svenskans beskrivning 34* (pp. 215–228). Lund.
- Hult, A. K. (2014). The Authentic Voices of Dictionary Users Viewing Comments on an Online Learner's Dictionary Before and After Revision. In A. Abel, C. Vettori & N. Ralli (Eds.), *Proceedings of the XVI EURALEX International Congress: The User in Focus* (pp. 237–247). Bolzano/Bozen: Institute for Specialised Communication and Multilingualism.
- Hult, A. K. (2016). *Ordboksanvändning på nätet. En undersökning av användningen av Lexins svenska lexikon*. (GNS 27.) Göteborg. Retrieved from https://gupea.ub.gu.se/handle/2077/49441
- Kann, V. (2010). KTHs morfologiska och lexikografiska verktyg och resurser. *LexicoNordica*, 17, 99–117.
- Kann, V., & Rosell, M. (2005). Free Construction of a Free Swedish Dictionary of Synonyms. *NoDaLiDa*, 2005, 1–6.

- Kokkinakis, D., Toporowska Gronostaj, M., & Warmenius, K. (2000). Annotating, disambiguating and automatically extending the coverage of the Swedish SIMPLE lexicon. *Proceedings of the Second International Conference on Language Resources and Evaluation*.
- Lyngfelt, B., Bäckström, L., Borin, L., Ehrlemark, E., & Rydstedt, R. (2018). Constructicography at work: Theory meets practice in the Swedish Constructicon. In B. Lyngfelt, L. Borin, K. Ohara & T. T.Torrent (Eds.), Constructicography. Construction development accross languages (pp. 41–106). Amsterdam: John Benjamins.
- Malmgren, S.-G. (2009): On production-oriented information in Swedish monolingual defining dictionaries. In S. Nielsen & S. Tarp (Eds.), *Lexicography in the 21st Century. In honour of Henning Bergenholt* (pp 93–102). Amsterdam: John Benjamins.
- Malmgren, S.-G. (2014): Svenska Akademiens ordlista genom 140 år: mot fjortonde upplagan. *LexicoNordica 21*, 81–98.
- Malmgren, S.-G., & Sköldberg, E. (2013). Scandinavian Lexicography. The Lexicography of Swedish and other Scandinavian Languages. *International Journal of Lexicography* 26(2), 117–134.
- Meyer, C. M., & Gurevych, I. (2012). Wiktionary: a new rival for expert-built lexicons? Exploring the possibilities of collaborative lexicography. In S. Granger & M. Paquot (Eds.), *Electronic Lexicography* (pp. 259–291). Oxford: Oxford University Press.
- Petzell, E. M. (2017). Svenska Akademiens ordbok på nätet. *LexicoNordica*, 24, 267–276.
- Rosqvist, B. (2014). Hårt arbete och sträng vila. *Svenska kollokationer i lexikografisk och lexikologisk belysning*. (Meijerbergs arkiv för svensk ordforskning 40). Gothenburg.
- Sköldberg, E., & Mattsson C. (2016). Ful- och finlexikografi? Om ordboksverksamhet i Sverige i dag och i morgon. In Hannesdóttir, A. (Ed.), *Framtidens lexikografi. Rapport från ett symposium i Göteborg:* 111–129. Gothenburg.
- Sköldberg, E., & Wenner, L. (2017). Folkmun och Luxikon. En jämförelse mellan två användargenererade ordbokssajter. In *Språk och stil 27*, pp. 21–48.

- Sköldberg, E., & Wenner, L. (2018). Amatörlexikografiska insatser på sajten Folkmun.se. In Á. Svavarsdóttir, H. Jónsdóttir, H. Hilmisdóttir, Th. Úlfarsdóttir (Eds.), *Nordiske Studier i Leksikografi 14*, pp. 237–245.
- *Språk- och folkminnen* (2017). Retrieved from http://www.sprakochfolkminnen.se/sprak/dialekter.html/
- Wolfer, S., & Müller-Spitzer C. (2016). How Many People Constitute a Crowd and What Do They Do? Quantitative Analyses of Revisions in the English and German Wiktionary Editions. *Lexikos*, *26*, 347–371.

STANJE ENOJEZIČNE LEKSIKOGRAFIJE: ŠVEDSKA

Prispevek podaja pregled švedske enojezične leksikografije. Omenjeni so glavni deležniki, tako komercialni kot nekomercialni, skupaj z opisom slovarskih izdelkov, ki jih ponujajo. Pregled izpostavlja, da je med švedskimi uporabniki slovarjev jasna in očitna tendenca, da se obračajo od tiskanih slovarjev k spletnim portalom in aplikacijam, kar močno vpliva na prakse komercialnih založba, kot so Norstedts, Bonniers in Natur & Kultur. Med vodilnimi nekomercialnimi ponudnik izpostavljamo Švedsko akademijo, Švedsko jezikovno banko in Inštitut za jezik in folklore. Švedska enojezična leksikografija pa ponuja tudi slovarje, ki jih izdelujejo nestrokovnjaki (tj. z uporabo množičenja).

Ključne besede: švedska enojezična leksikografija, švedske leksikografske ustanove, švedski spletni leksikografski portal, švedski uporabniški slovarji



To delo je ponujeno pod licenco Creative Commons: Priznanje avtorstva-Deljenje pod enakimi pogoji 4.0 Mednarodna. / This work is licensed under the Creative Commons Attribution-Share-Alike 4.0 International.

https://creativecommons.org/licenses/by-sa/4.o/