

The Saga of *Norsk Ordbok*: A scholarly dictionary for the Norwegian vernacular and the Nynorsk written language

Oddrun Grønvik

The 9th of March 2016 saw the launch of *Norsk Ordbok*, a twelve-volume scholarly dictionary of the Norwegian vernacular and the Nynorsk standard language. *Norsk Ordbok* fills twelve volumes of 9,600 pages, has about 11 million words of text, holds 330,000 entries and ca 15,000 fixed phrases. It took 86 years to complete since the material collecting started and until volume 12 was out. Two thirds of the editing happened after 2000. The dictionary as well as much of the evidence (contained in the Norwegian Language Collections, cf. Grønvik 2020) is freely available on the web (<http://www.norskordbok.uio.no>).

The full story of a twelve-volume scholarly dictionary could easily fill another volume, but in this article only a few points will be addressed, i.e. (1) the linguistic backdrop, (2) the dictionary project and its source material, (3) the digitisation project NO2014, and (4) the future.

1. The linguistic and historical backdrop to *Norsk Ordbok*

Norway has a broken history; independence until the end of the 14th century, subordination under Denmark until 1814 and under Sweden from 1814 to 1905, and independence again since 1905. These political changes have had a profound influence on the Norwegian language, which in turn has affected the formation of written standards and the scholarly lexicography for Norwegian. The chief result is that Norway today has two written standards, Bokmål and Nynorsk, which are close cognates, and which are each documented in a major dictionary. *Norsk Ordbok* documents the Nynorsk written standard and all Norwegian dialects.

The historical background can be summarised as follows (cf. Haugen 1976; Vikør 1995 p. 51 ff. and 92 ff.):

The spoken languages of medieval Norway, Iceland, Sweden and Denmark must have been mutually comprehensible, but resulted in different written practices. In this period, the written language of Norway was what is now called Old Norse.

Once the administration of Norway was transferred to Denmark, Old Norse was gradually replaced by Danish, until by the end of the sixteenth century Danish became the language for civic administration. Norway was not allowed a university until 1811, so tertiary education meant a

long stay in Copenhagen. Danish was the dominant written language in Norway until after 1905.

In the same period, the Norwegian vernacular changed so much that a revival of Old Norse as a written standard after 1814 was unthinkable – the spoken dialects and the old written standard were too far apart.

The 1814 Constitution states that legislation should take place in the Norwegian language, but this was a shield against a Swedish takeover. The choice of Danish as an administrative language nevertheless left Norway with a national legitimacy problem – the idea of a separate Norwegian national identity, and the need for an independent state, was questioned. The language issue became the question of the day from the mid-19th century, and two solutions were presented, though not shaped into opposing camps until the end of the nineteenth century.

The response to the legitimacy issue was initiated by members of the Royal Norwegian Society of Sciences and Letters (DKNVS). The society looked actively for someone who could document the Norwegian vernacular language, and prove (a) its connection to Old Norse, and (b) its separateness from Danish and Swedish. Because of the diglossic situation - Norwegian and Danish were close cognates, and Danish spoken in Norway was phonologically adapted to Norwegian - the difficulty was finding a trained linguist who was close enough to ordinary people to gather trustworthy linguistic information. The problem was solved when the self-taught linguist and lexicographer Ivar Aasen (1813–1896) presented himself for the task. Aasen was funded from 1840 onwards and throughout his lifetime, first by DKNVS, then by Stortinget. Within his lifetime, Aasen documented the grammatical structure and the lexicon of Norwegian in a series of works culminating in *Norsk Grammatik* (1864) and the dictionary *Norsk Ordbog med Dansk Forklaring* (1873). The orthography expressed in the headwords of Aasen's 1873 dictionary was also his proposed standard for a common, wholly Norwegian written standard, the forerunner to today's Nynorsk (New Norwegian). Aasen's work put an end to the legitimacy doubts – Norwegian was a



Oddrun Grønvik obtained a B.A. with honors in English Literature and Language from St. Hilda's College, Oxford, has the degree Candidate of Philology at University of Oslo, and was in 2006 granted an honorary PhD from the University of Zimbabwe for her work on dictionaries for African languages. She has worked in the publishing industry (1970-1979) and as consultant at the Norwegian Language Council (1979-1987), and in 1987 became Assistant Professor at the University of Oslo and joined the editorial staff at *Norsk Ordbok*, serving as Chief Editor between 2006-2016 with special responsibility for lexicographical and digital development and training. In addition she has served as academic project manager for the NORAD-funded African Languages Lexical Project, and on the standardisation committee of the Norwegian Language Council. Dr Grønvik has written widely on the Norwegian language, language development and standardisation, lexicography and digitisation of language resources.
oddrungronvik@iln.uio.no

Ivar Aasen:
Norsk Grammatik
§ 384 (1864)

"What especially concerns the word forms, is that the particularities belonging to each of them are so many, that they cannot easily be dealt with otherwhere than in a dictionary."



separate West-Nordic language, descended from Old Norse, while modern Danish and Swedish stem from East-Nordic.

Aasen's work was made possible by the development of the comparative methodology of 18th and 19th century historical philology. He systematically documented the Norwegian dialects, employed the comparative method to establish a common pattern for phonology, morphology, lexicon and syntax, using Old Norse as a touchstone, but including nothing that was undocumented in his time. In shaping his proposed standard language he also took the standards of Swedish and Danish into account, to avoid unnecessary differentiation from what people were used to seeing in print.

To the ruling classes of Norway, however, the idea of even trying to establish a wholly Norwegian written standard, for everyday use in competition with Danish, seemed ridiculous and unthinkable. This would mean giving cultural hegemony to an uneducated, though literate, country population. At the same time, something had to be done to nationalise the Norwegian version of Danish, clearly diverging from the Danish of Denmark. The counter-solution to Nynorsk favored adapting standard Danish to Norwegian phonology and including typically Norwegian words in the lexicon. A gradual transition from a Danish to a Norwegian written standard, based on "educated everyday speech" was envisaged. The first orthographic reform of Danish in Norway came in 1907 and established the forerunner of today's Bokmål.

At the end of a long and fierce political struggle, the Norwegian parliament in 1885 voted to give both standard languages official standing as languages of instruction

and leave it to each school board to choose which one to adopt. An earlier parliamentary decision, in 1874, had tasked teachers with adapting their oral instruction in class to the dialect of their pupils, instead of the other way round. Since then, Norwegian has been expressed in two written languages. Since 1929 these have been termed Nynorsk and Bokmål (Vikør 1995; Hovdhaugen 2000).

2 The *Norsk Ordbok* dictionary project and its source material

Plans exist back to 1911 for scholarly lexicography for Norwegian, in the form of committee reports and applications for funding. From 1920 onwards some funding was achieved for starting language collections — slip archives with indexed excerpts, according to the best practice of the times. The resulting lexicographical work was envisaged both as a joint presentation of all spoken and written Norwegian, and as a separate work for each of the standards. Separate scholarly dictionaries became the solution and resulted in the parallel projects of *Norsk Riksmålsordbok* and *Norsk Ordbok - Ordbok for det norske folkemålet og det nynorske skriftmålet*. *Norsk Riksmålsordbok* was published in four volumes 1928-1958, with a supplement in two volumes published in 1995.

The split into two projects had an ideological basis with inevitable practical implications. The scholarly dictionary for the Danish-derived standard was to be based on printed literature of Norwegian authorship from 1814 onwards, supplied by speech materials representing "educated everyday speech". The scholarly dictionary for Nynorsk was to be based on the Norwegian vernacular in all its varieties, as documented back to about 1600, and on Nynorsk literature, which came into being from the late 19th century onwards. The first project regarded speech as a supplementary category and a dialect label as a warning; the other one saw the dialect materials as primary source material, expanded and developed through literary use. Sources as well as the lexicographical treatment of them were to be too different for the projects to be compatible within one framework.

Norsk Ordbok got off to a belated start in 1930. A grand plan for material collection, with a small editorial staff supported by volunteers, was drawn up and drew a gratifying response: 600-700 volunteers came forward within a year or two, and by 1940 the collections encompassed one million slips, 20 percent documenting speech, the rest drawn from written sources. At the same time, a rough first version was drafted on the basis of existing Nynorsk dictionaries and some large dialect

collections. This draft manuscript held 130,000 entries and covered 13,500 pages of typescript. The plan was to expand this manuscript with the materials from the language collections and end with a modern dictionary of 4-5 volumes.

Work on *Norsk Ordbok* was halted during the second World War, and started again in 1946. A review period led to the following three decisions: (a) continue collecting, especially oral materials; (b) draw up a detailed plan for the dictionary microstructure, so as to do justice especially to the richness of the sources; (c) start editing all over again, focussing on full use of the Language Collections with the draft manuscript as a guideline. On the basis of this very ambitious plan, the first fascicle was completed in 1950 and the first volume in 1966. At this time, the completed dictionary was thought to reach eight volumes at the most.

The group of editors increased slowly. When I was recruited in 1987, I became the eighth editor and the second woman editor. At that time two volumes were out and the third completed in manuscript. The Language Collections had quadrupled in size and the alphabet progress had slowed down. A little arithmetic showed that if work continued at the rate then current, *Norsk Ordbok* would be completed around 2060 and reach 16-20 volumes — a plan which was unlikely to get funding. These facts were therefore kept quiet.

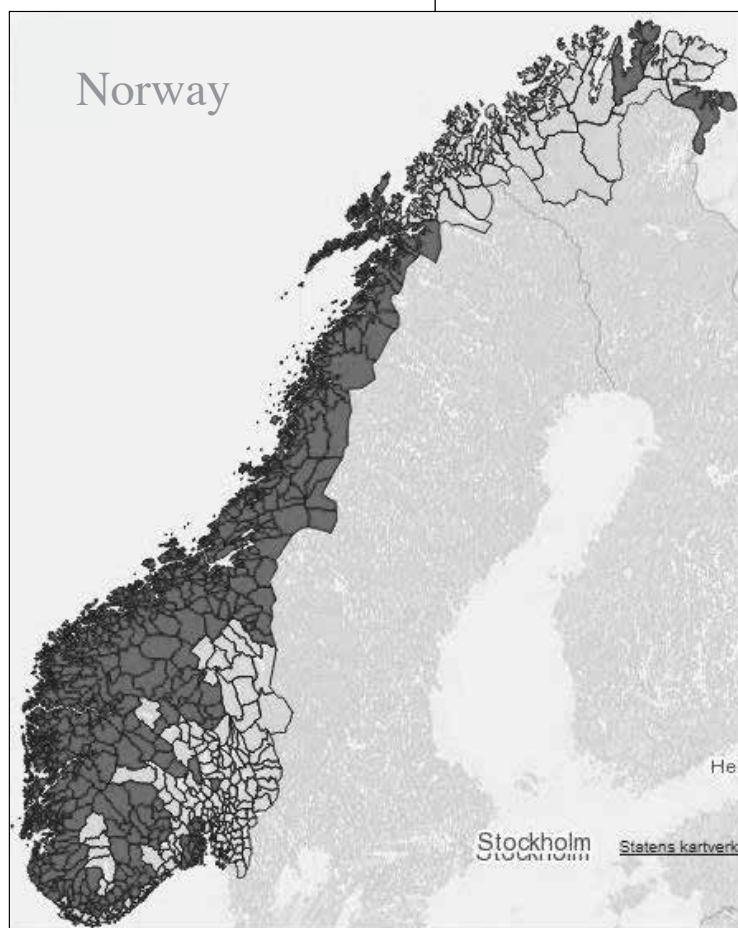
Norsk Ordbok needed a miracle, and the miracle turned up in the shape of a huge digitisation project for the university collections of the whole of Norway, designed to counteract nationwide unemployment when the Norwegian telephone and telegraph services went digital. Through the Documentation Project (1991-1997), run by Christian-Emil Smith Ore at the University of Oslo, key components of the national Language Collections were stored in databases and on the Web by 1997, i.e. the excerpt archives, the draft manuscript of 1940 and a number of other resources. All components were then coordinated in a digital index – the Meta Dictionary – with base forms and part of speech as in *Norsk Ordbok*. The Meta Dictionary at present holds about 550, 000 entries for Nynorsk.

3 The Digitisation Project NO2014 (2001–2016)

As the millenium approached, Norwegian authorities were planning another jubilee — the bicentenary of the Norwegian constitution in 2014. *Norsk Ordbok* was chosen as one of the bicentenary projects, on the basis of a carefully worked out

production plan. The basis for this plan was the conviction that trained linguists could become efficient scholarly lexicographers within one year, and after that meet production deadlines as planned. In order to succeed, *Norsk Ordbok* would need roughly 265 man years of efficient lexicography within 14 years — 2001 to 2014. We thought we could do that, given (a) competent and tough management, (b) scholarly computer developers, (c) enough linguists, and (d) funding. All of these factors were equally important, and the planned project, named NO2014, would be doing a tightrope act from beginning to end. It was worth trying.

The conviction that scholarly lexicographers could be trained quickly and efficiently ran counter to traditional views of training needs for scholarly lexicographers. When I started in 1987, the general assumption was that training as a scholarly lexicographer would take at least five years, and the time would be spent in getting to know the collections, mastering a multitude of conventions, and accepting the need for extensive crosschecking and proofreading. Previous experience as a linguist would certainly be utilised, developed and challenged in handling very complex materials, but language analysis was only one of many tasks, and they all



The screenshot shows the online version of the Norwegian dictionary 'Norsk Ordbok'. The search bar contains the word 'støl'. The search results show the entry for 'I støl substantiv, hankjønn'. The entry includes a definition: '1. stad nytta som mjølkeplass i avlees hamnehage med sei (IV,1) og beitemarker ikring (der buskapan, seri kyr og geiter, er om sumaren); plass i utmarka med stølshus; sæter (I) (Lom1743, Va1743, Land1743, Elker, Modum og Sigdal1744, Tinn1743, Nhl1750, W, Seljord1786, V50, H, Snn-1810, VTel1821, C, A, R; Heimfestingar: Kvikne, Gbr, Slidre, Lunner, Hall, Adal, Vestf mst, ØvTel, IAgd, Vestl, Tr sumst, No sumst, Lebesby, Gamvik, SVaranger Kjøp. Litterære kjelder: ReintonTSR 23f, Østb. VI,23)'. Below the definition are several usage examples with their literary sources: 'fara, flytja, reisa til, på stølen vera på stølen om sumaren', 'til alle gardar og bruk var det ein eller fleire stølar (Heimfesting: Hemsedal Kjørt)', 'dei henta mjølk åt oss ifrå stødien (Litterær kjelde: jfr KrohnFerd 17)', 'alle gar'e sætre på di stuiom i høgjeille, som dam sit på den dag idag (Litterær kjelde: Kl.R 100)', 'han køyrde høy frå støylane (Litterær kjelde: Visl.GG 105)', and 'flytta stølar flytta frå støl til støl (Heimfesting: Ryf Kjørt; Litterær kjelde: ReintonSr I,449)'. There is also a section for 'i samb til støls' with examples like 'opp til sætra; på sætra', 'koma til støls', 'gå til støls driva ein støl (Litterær kjelde: GloppenFitje)', and 'dei låg til støls med fenaden (Litterær kjelde: JansonMS 53)'. A 'Map button' is located on the right side of the entry content.

The entry *støl* in the online version of *Norsk Ordbok*

seemed to hold equal significance.

Norsk Ordbok was one of many projects to embrace the computer at the end of the 20th century. The decisive experience for us on the issue of training enough lexicographers in a fairly short time, was participation in the ALLEX Project (1991-2006), a Norwegian and Swedish funded research project designed to provide monolingual dictionaries for the African languages of Zimbabwe. The ALLEX Project proved that mother-tongue linguists could become efficient scholarly lexicographers in a very short time, working through a lexicographical interface based on an analysis of the relevant language, storing results in databases, dealing with oral materials in corpora, etc.

Categorising and commenting on language through well worked-out software is not only a tool for efficiency, it is also an immensely effective tool for learning and mastery. This conviction combined with the assurance of computerisation support from EDD (the Unit for Digital Documentation at the University of Oslo), covered points (b) and (c) above. The Norwegian Parliament guaranteed point (d), with funding through the Ministry of Culture and the University of Oslo. NO2014 also had the immense good fortune to attract two directors¹, one after the other, who had all the qualifications one

could wish for in professional and human terms. The positions as chief editors – a group of four – were held by former staff members. Tasks were allocated according to project needs. I was made responsible for digitalisation and training, and this account is naturally coloured by my particular experience.

A premise for funding was moving the entire project to a digital platform. We took that to mean not only producing the dictionary itself, but also being able to access and sort digitalised materials, take care of the sorting, make sure that no entry lacked materials, and saving the finished product in a form that allowed different types of presentation of the finished product (Grønvik 2005).

An important decision concerning the software structure was to make a maximum format the standard, always allowing for the most extensive and complex entry, rather than having a more restricted basic format which might have to be extended. The standard sense unit therefore caters for definition, usage examples, sub-definitions with usage examples, multiword expressions with several senses, and finally compounds in which the entry headword appears as the initial or the final part, plus of course source tables for literature and geographical location.

The editorial interface was the first thing to get finished. By 2013, *Norsk Ordbok* in digital form was contained within one application which was able to (1) generate

1. The Project directors of *Norsk Ordbok* 2014 were Kristin Bakken 2002-2008 and Åse Wetås 2008-2015.

entries from indexed materials, with a link to the materials, (2) provide a tool for analysing linked materials and storing the analysis, (3) generate the entry head (the identifying information of the entry) from a separate full form register, (4) present a form where the edited text can be linked directly to the materials underlying each definition or description, (5) allow supervision of production flow at the micro and macro levels, (6) present the finished product in an optimally accessible fashion (paper: pdf with preset style sheet corresponding to the print typeface; web: settings for web presentation on different reading tools), and (7) provide a format for longtime storage of the product linked to its sources. The software package became a sort of lexicographical factory, designed to allow editors to concentrate on analysing and editing (Grønvik and Ore 2013).

The *Norsk Ordbok* software is now used in three other dictionaries, *Bokmålsordboka* and *Nynorskordboka* being the best known (see <http://ordbok.uib.no/>, cf. <http://dictionaryportal.eu>).

When the project NO2014 started in 2001, the alphabet stretch a-h was already edited, with great care and consistency, and deep respect for the materials contributed over the years, especially the oral materials. A primary task in the digitalisation process was to take care of what can be termed best practice in the pre-digital editorial work. Two tasks stood out: (a) the careful identification of formerly unstandardised dialect word forms, and finding for them a standard form consistent with modern Nynorsk orthography; (b) the treatment of multiword expressions (MWEs), which proved to have been a major difficulty in the pre-digital entry schema. The first task became a permanent concern for the project management, especially in offering all new editors training in handling Norwegian and Nordic dialectology, synchronically and diachronically, but also in giving particular attention to the standardising of new dialect materials which were added to the Language Collections during the project period. For the second task, training in identifying MWEs was offered on a permanent basis, but we also created a software template for the registration and editorial handling of MWEs, making them directly searchable.

Identifying both poorly documented word forms and MWEs was greatly helped by important additions to the digital Language Collections. The most important items were a corpus for Nynorsk literature covering the period 1866–2010, now at 105 million tokens (Nynorskorpuset), and the digitalization of 65 dialect dictionaries

(Norway has more than 400 dialects) in a common portal allowing cross-searches in headwords.

A Web edition of *Norsk Ordbok* was launched in 2012 (<http://norskordbok.uio.no/>), showing the section of the dictionary edited and completed in the relational database, today from “i” to “å” in the alphabet. From January 2014, the Web edition has been linked to a digital map of Norway, and thus been able to show geographical usage extent for word forms, senses and expressions. This edition is popular, and so is the map function.

When the NO2014 project was nearing completion, we also published our editorial handbook through the NO2014 website (*Redigeringshandboka* 2016). The project parole throughout was to encourage users to look behind the edited text into the materials of the Language Collections, raise questions and demand response. In public interaction, publishing the guidelines, which have the role of a method chapter in a dissertation, has turned out to be useful.

When *Norsk Ordbok* was completed, more than 200,000 headwords that had never been lexicographically treated, had an entry in a scholarly dictionary, while the central vocabulary of Norwegian had received in-depth treatment on the basis of written and oral materials covering the whole country and four centuries of documentation (Grønvik 2017).

A fully sourced account of the history of *Norsk Ordbok* (in Norwegian) will be found in the Festschrift published together with the final volume (Karlsen et al. 2016).



Launch of *Norsk Ordbok*, 9 March 2016



Series of fully bilingual dictionaries for Nordic and major European languages.

DANSK

Danish-English
/ English-Danish
Danish-French
/ French-Danish
Danish-German
/ German-Danish
Danish-Spanish
/ Spanish-Danish

NORSK

Norwegian-English
/ English-Norwegian
Norwegian-French
/ French-Norwegian
Norwegian-German
/ German-Norwegian
Norwegian-Spanish
/ Spanish-Norwegian

SVENSK

Swedish-English
/ English-Swedish
Swedish-French
/ French-Swedish
Swedish-German
/ German-Swedish
Swedish-Spanish
/ Spanish-Swedish

Based on the Global Series and other resources and developed by K Dictionaries 2017.

4 From the University of Oslo to a new life at the University of Bergen

In June 2014, the University of Oslo decided to end its commitment to Norwegian lexicography, and get rid of the Language Collections, which comprise archives going back to the 1880's and covering far more than the Nynorsk and dialect sections. NO2014 was half way through editing volume 12 when the project staff was sacked. Despite great difficulties, *Norsk Ordbok* did get finished in good order, but there was a delay of more than a year; volume 12 was sent to the publisher, Det Norske Samlaget, on November 24, 2015, and the launch came in March 2016.

By that time, the Norwegian government, through the Ministries of Education and Culture, had decided that the Language Collections, with the dictionaries *Norsk Ordbok*, *Bokmålsordboka* and *Nynorskordboka*, represent essential linguistic infrastructure, and therefore were too important to be left to the management of the University of Oslo alone.

Provided that funding was allocated, the University of Bergen had volunteered to house the Language Collections (comprising collections for Bokmål, Nynorsk, Old Norse and place names). After inventorying in the winter of 2015-2016, more than 70 tons of books and archives were moved in the summer of 2016. The transfer of the digital collections started about the same time, the first components being run from Bergen from September 2016. The total move is a very extensive operation still in process, involving recruiting and training of research, ICT and administrative staff. In February this year an application for revision and full digitisation of *Norsk Ordbok* a-h was submitted by the University of Bergen to the Ministry of Culture, and it was – so far and fingers crossed – well received (for revision plans, see Berg-Olsen et al. 2015).

This is how matters stand.

The cost of completing *Norsk Ordbok* through the Project NO2014 (2001-2016) stands at 260 million NOK, somewhere around 27.5 million Euro. This sounds like a lot of money, though it wouldn't buy many kilometers of road. However, it is enough not to be thrown away lightly, especially when there is visible and vocal public support for maintaining both the Language Collections and the dictionaries. In the future, Norwegian lexicographers will have to continue to serve the public, both in Norway and internationally, through developing Norwegian lexicography as best they can. At least we now know that we can do it!

References

- Berg-Olsen, S. and Wetås, Å. 2015.** Revision and digitisation of the early volumes of *Norsk Ordbok*: Lexicographical challenges. In Abel, A., Vettori, C., and Ralli, N. (eds.), *Proceedings of the 16th EURALEX International Congress*. Bolzano. EURAC research, 1075-1086.
- Grønvik, O. 2005.** *Norsk Ordbok* 2014 from manuscript to database – Standard Gains and Growing Pains. In Ferenc, K., Kiss, G., and Pajzs, J. (eds.), *Papers in Computational Lexicography Complex 2005*. Linguistics Institute, Hungarian Academy of Science. 60-70
- Grønvik, O. 2020.** The lexicography of Norwegian. In Hanks, P. and Schryver, G.-M. (eds.), *International Handbook of Modern Lexis and Lexicography*. Berlin and Heidelberg: Springer-Verlag. Forthcoming.
- Grønvik, O. and Ore, C.-E. S. 2013.** What should the electronic dictionary do for you – and how? In Kosem, I. et al. (eds.), *Electronic lexicography in the 21st century: thinking outside the paper. Proceedings of the eLex 2013 conference, 17-19 October 2013, Tallinn, Estonia*. Eesti Keele Instituut, 243-260.
- Haugen, E. I. 1976.** *The Scandinavian languages: An introduction to their history*. Cambridge, MA: Harvard University Press.
- Hovdhaugen, E. 2000.** Normative studies in the Scandinavian countries. In Auroux, S. et al. (eds.), *History of the Language Sciences*. Berlin and New York: Walter de Gruyter I-II, 888-893
- Karlsen, H. U., Vikør, L. S., and Wetås, Å. 2016.** *Livet er øve, og evig er ordet*. Oslo: Det Norske Samlaget.
- Vikør, L. S. 1995.** *The Nordic languages. their status and interrelations, 2nd ed.* Oslo: Novus.

URLs

- Bokmålsordboka* and *Nynorskordboka*: <http://ordbok.uib.no>
- The Documentation Project (1991-1997): http://dokpro.uio.no/engelsk/drawer_to_screen1.html
- ENeL Dictionary Portal: <http://dictionaryportal.eu/en/>
- Norsk Ordbok*: <http://no2014.uio.no/perl/ordbok/no2014.cgi>
- Nynorskorpuset*: http://no2014.uio.no/korpuset/conc_enkeltok.htm
- Redigeringshandboka 2016*: <http://no2014.uio.no/eNo/tekst/redigeringshandboka/redigeringshandboka.pdf>