

Turning a Dictionary Around – Norsk Ordbok 2014

Dr Kristin Bakken

Director of Norsk Ordbok 2014

University of Oslo

P. Box 1021 Blindern

0315 Oslo

Norway

kristin.bakken@inl.uio.no

Abstract

The author presents the new project Norsk Ordbok 2014 which has been organized around the old scientific dictionary *Norsk Ordbok*, one of two national dictionaries in Norway. Since 2002 the dictionary has been reorganized, and the paper presents the new management, the work done to computerize the editorial routines and the challenges connected to the training of new editors. In order to meet new strict deadlines, the management of the dictionary has been strengthened, thus enabling much closer planning. The computerization scheme is ambitious, and the ultimate goal is fully integrated digital editing routines. Archives, corpus and to some extent also literature are accessible on screen, and the analysis of relevant material is done on a digital desktop. Finally, the dictionary entry is written into an input form where preset menus help the editors getting the structure and formal layout of the article right. The paper concludes by discussing the effects of the reorganization on the dictionary itself.

1. Background

Norsk Ordbok is one of two scientific national Norwegian dictionaries. Norway has two official written standards, *bokmål* (literally “Book language”) and *nynorsk* “New Norwegian”). *Bokmål* is the written standard favored by the majority of Norwegians, it is used by most of society’s elite, by people in the biggest cities and it is the language of our classical authors Ibsen, Hamsun and Undset. *Bokmål* is covered by the dictionary *Norsk riksmålsordbok*, a three-plus-two volume dictionary consisting of some 7000 pages.

When *Norsk Ordbok* was conceived in 1930, it was against this background. The New Norwegian community wanted to raise the prestige of their written standard. They wanted to demonstrate the full range of possible domains in which this standard was used, and maybe most important, they wanted a conscious effort in order to extend the New Norwegian lexicon. *Norsk Ordbok* was therefore planned as a combined dictionary, both covering the dialects and the written language. New Norwegian was constructed on the basis of the dialects, and ever since it has been fed by new dialect words. In *Norsk Ordbok* all Norwegian dialect words are explicitly related to the New Norwegian written standard, and thus given literary status. It is significant that this dialect focus greatly increases the dictionary’s national relevance to many *Bokmål* users.

2. Norsk Ordbok by 2000

By 2000 the dictionary consisted of three volumes covering the alphabet from *A* to *Gi*-. The fourth appeared in 2002. In the first decades after 1930, the resources were largely spent accumulating a dictionary archive. In 2000 this archive contained some 3,2 million excerpts.

They covered New Norwegian literature on the one hand, and orally collected dialect material on the other. The staff consisted of seven editors and two assistants. The alphabet progress was slow, and was getting slower by every volume, so that the size of the finished dictionary was continuously growing, and the finishing year was consequently ever moving farther down the 21st century. The archives had been computerized during the 1990s, but the challenge now lay on the editors to start exploiting the opportunities that such computerization entailed.

3. Norsk Ordbok 2014

In the years 2000-2002 there was a political initiative carried out to turn the dictionary around. On the condition that the dictionary was completed by 2014, the government granted the money to expand the staff and to develop new time-efficient computer based routines. The dictionary was thus reorganized as a project called *Norsk Ordbok 2014*. The management side of the project was substantially strengthened. Whereas the editors formerly by and large managed themselves, the new project now has a director and an external board. The technology challenges were met by allocating two full positions to computer development. And the new funding allowed us to employ 10 new editors the first year. The project was presented with two conditions: to finish the dictionary by 2014, and to confine the dictionary to 12 volumes.

4. The reorganization process

The rest of this paper will be spent focusing on the reorganization process and some of the challenges that we have had to deal with since we started afresh in June 2002. Although we had the funding, the people and the new guiding conditions, no map was drawn for us as to how we were to meet the challenges associated with almost all aspects of our work.

4.1 Management

The director was given both administrative and editorial responsibilities. The latter did not pertain to the manuscripts themselves, but to editorial principles and priorities. Experience has proven this double responsibility invaluable. Another model of management was rejected early in the process, one could have divided the responsibilities between a head editor and a managing director with purely administrative powers. Such a solution would probably have been more ideal for an established dictionary project. But in a situation where both the administrative and the editorial routines were to be revised, and as it turned out, also the editorial principles themselves, it has been essential to have the ultimate responsibility for all these interdependent aspects of the dictionary allocated to one person.

Editors of a scientific dictionary are highly competent researchers. Our staff had been with the dictionary for 10 to 40 years when the project got under way in June 2002. Starting up, one of the director's main challenges was to see to that all the valuable experience the staff possessed, was channelled in a constructive way into the revision process. Examples are work with the specifications of the editorial program, the design of the new text corpus, the elements in the training program, the principles for planning lemma selection, the space allotted to individual words and alphabet chunks, and the chronological and sequential

progression of the dictionary. We solved most of these issues by forming small ad hoc work-groups during the first year. Now we have established a more permanent advisory group working closely with the director in all matters of lexicographical substance. By including the established editors in the reorganization process, we have managed to build an essential bridge between the “old” dictionary and the new project.

4.2 Computer development

Computerization was seen as a necessary prerequisite of turning *Norsk Ordbok* around. One basis was the work that The Documentation Project (now DOK) had done during the 1990s, when the dictionary archive was computerized. As a result of this effort, all index cards had been made accessible in facsimile on screen, and a computerized index enabled the users to perform searches. The challenge to integrate the use of this digital archive in the editorial routines remained, however.

Unit for digital documentation (DOK) at the University of Oslo had by the time the project came into being, started working on two computer applications that aimed at a new, fully integrated computerized editing process. The goal of this work was twofold. First, it was important to make the editing process more efficient in order to speed up the dictionary production. Secondly, there was a strong qualitative motivation for the initiatives taken. The idea was to help the editors do more consistent work, and improve the empirical foundation on which the articles were written. The two applications that had been conceptualized before the project was officially started, was on one hand an editing program, and on the other hand an extension of the digital index so that *all* computerized sources of our dictionary were linked in a common database.

The index was called *Metaordboka* or “the Meta-dictionary” (cf. Ore 1999; Svardal 2003), and it marks an important intermediate level between the archives and the finished dictionary. In this database we now have all computerized sources linked to an all-comprehensive list of the dictionary’s lemmas. It is accessible to the public over the internet (cf <http://no2014.uio.no>), thus enabling our users to look up words not yet edited by our staff. It has proven very valuable for our dictionary planning, for it gives us all the information we need as to how many token of each lemma that are registered in our archives, what sources they appear in, and how many lemmas we have to consider in all. It also gives us the basis to make an automatic lemma selection application, and the numbers that are necessary to plan the size of each article.

The editing program has been designed as an integrated extension of the Meta-dictionary. The rather complex structure of the articles in the dictionary has been formalized, so that that the program takes the editor through an input form, helping him along the way with predefined menus, and highlighting the semantic hierarchy of the article. The program reduces the time needed for manuscript corrections, as typography, abbreviation standards, and some terminology are given by the program. Examples of preset menus are comprehensive lists of bibliographical references, geographical references and of grammatical and etymological classifications. Equally important from a time-saving perspective, is the fact that the editing program greatly reduces the training time for new editors since the program leads a new editor by the hand through the highly formalized structure of a *Norsk Ordbok* entry. Both the Meta-dictionary and the editing program were

primarily planned as measures to speed up production time, but both applications actually also improve the *quality* of the dictionary, especially when it comes to consistency.

The third major initiative toward the goal of completely computerized editing routines, is our new corpus (cf. <http://no2014.uio.no>). It was put together during the first year of our project's existence, and marks an important extension of *Norsk Ordbok's* literary basis. By June 2004 it consists of some 30 million words, and Daniel Rødings who has developed it, has seen to that it is fully integrated with the editing program and the Meta-dictionary, thus enabling the editors to feed the digital archive with corpus citations and concordances, and to copy such citations directly into the proper slots in the editing program.

One of the worries of the old editors was that it would be difficult to sort the relevant material on a digital desktop. They proved to be partly right, but only as to the size of the desktop. The sorting is now about to be done within the limits of the Meta-dictionary, and the space problem has been met by extending the digital desktop to two screens.

4.3 Training

The establishment of our new deadline, that is the year 2014, was based on a resource analysis that concluded with the need to employ and train 20 new editors in 3-4 years. This has really proved to be the biggest challenge for *Norsk Ordbok 2014*. Our project has turned out to be an attractive place of employment for young linguists, and we have not had any problems attracting highly skilled editors to be. We had seven experienced editors when the project started in 2002. By 2003 we have trained ten new ones. It takes at least a year to train a new editor to be able to meet the production goals and deadlines, and new editors need teaching, tutoring and close manuscript surveillance. By the end of 2003 we see that the established editorial staff is a bit too small to be able to absorb and train the number of editors we now need, at least if the old staff is expected to keep up its own manuscript production. A beneficial circumstance is of course the editing program, which in some respects move the teaching away from the tutors and over to the program which has tutorial functions. Nevertheless, we may be forced to extend the recruiting phase simply to be able to train sufficient many editors in a proper way.

5. Has the reorganization process affected the dictionary itself in any way?

In an earlier paper I posed the question as to if, and how the new project has affected the dictionary form or content (Bakken in press). The starting point of the new project was to complete the dictionary as one knew it by the year 2014. In addition there is now a 12 volume limit on the work. Although there was no initial ambition to change the dictionary as such, we now see that it *will* be affected by the conditions that the project poses on it.

First, our empirical basis has been greatly expanded by the use of our new corpus. This expansion puts an emphasis on newer texts and on other genres compared to our old archive, and there is no doubt that we now have a better and fuller basis on which to found our dictionary entries. Secondly, the Meta-dictionary makes it less likely for any editor to miss major sources to a lemma. Thirdly, the editing program guarantees that our dictionary will be more consistent than it used to be before. In specifying the requisites that had to be met by the program, the old staff discovered surprisingly many inconsistencies among themselves and in the earlier texts. Such inconsistencies have now been harmonized by the

program frames. Finally, compared to the fourth volume of the dictionary, the entries in the next eight volumes will be less elaborate, especially as to the number of citations quoted. Besides, we need to put a bottom limit to the number of instantiations a lemma must have in our archive to qualify for a separate entry in the dictionary. Both these limitations may be viewed as a loss for *Norsk Ordbok* as a scientific dictionary. However, the Meta-dictionary secures the public access to sources and lemmas that have not made their way into the dictionary, thus making the line between the dictionary and its sources less clear-cut than before. And although some would view these new quantitative bounds as a loss to the dictionary as such, I have tried to show that the new project actually secures higher quality according to non-quantitative evaluation parameters.

References

- Bakken, K.** in press. 'Norsk Ordbok 2014 – Rammer og Utfordringar'. To appear in *Nordiske studier i leksikografi* 7. Rapport frå Konferanse om leksikografi i Norden, Volda 21.-23. mai 2003. Oslo: Nordisk forening for leksikografi.
- Ore, C.- E.** 1999. 'Metaordboken – Et rammeverk for Norsk Ordbok'. *Nordiska studier i leksikografi* 5. Rapport från Konferens om leksikografi i Norden, Göteborg 27-29 maj 1999. Göteborg: Nordiska föreningen för leksikografi.
- Svardal, T.** 2003. 'Norsk Ordbok – Om Normeringa av Metaordboka'. *Nordiske studier i leksikografi* 6. Rapport fra Konference om leksikografi i Norden, Tórshavn 21.-25. august 2001. Tórshavn: Nordisk forening for leksikografi.