

# The BRO-project, a bridge in the wild, Norwegian linguistic landscape

Ruth Vatvedt Fjeld & Petter Henriksen



Keywords: *national dictionary, xml, collaboration.*

## Abstract

The Norwegian language falls into two main variants - bokmål and nynorsk. The majority's variant is bokmål, used by over 90 % of the population. Historically, bokmål again falls into several sub-variants, but now the two main sub-variants- riksmål and bokmål proper - are practically united in one common norm. This norm is being documented in the national dictionary project bearing the symbolically significant name BRO ('bridge'). The article presents the background for the BRO collaboration, and sketches a concrete and feasible plan for the lexicographical documentation of the common norm. A challenge lies in the choice of lemma sign form and the presentation of bokmål's wide variety of optional forms, where also style nuances play a role. The same applies to the choice of examples and collocations and other multi-word lemmas. Both challenges arise from the need for freedom of expression within the norm, which is typical of Norwegians' preference to mark identity through language.

## 1. Introduction: Why build a linguistic bridge in Norway?

There are two official written varieties of the Norwegian language. As Norway is a large and rugged country, it also has many oral dialects. The Norwegian nation, reborn in 1814 after several hundred years of Danish rule, wished strongly to have its own national language. By and by the principle was established that every inhabitant should have the right to choose her or his own written variant, as close to one's oral dialect as possible – still an important credo in Norwegian language planning. As a result, the two written standards favor dialects from either Eastern or Western Norway. The eastern dialects are closest to "bokmål" (literally 'book language' in opposition to the spoken dialects), in English by some writers called 'Dano-Norwegian', due to the Norwegian language history. In this article we will call this variety Standard Norwegian, which is the absolute majority variant in use in the main parts of the country, area-wise as well as population-wise. On the other hand, "nynorsk", in English Neo-Norwegian, used by approximately 10 % of the population, above all reflects the dialects in the western parts of Norway. Most nations have their hands full compiling one national dictionary. Norway, with its 5 million inhabitants, consequently needs two. This is interesting seen from the vantage point of dictionaries and national identity, as these different standards mean splitting instead of unifying the population.

For the minority variant Neo-Norwegian, work has been going on since the 1930s creating a paper-based national dictionary bearing the name *Norsk Ordbok*. The label Norsk (Norwegian) is used as opposed to the Danish dictionaries from earlier times. This twelve volume dictionary is planned to be completed in 2014.

As for the majority's written variety of today, Standard Norwegian, there have traditionally been two subvarieties, *bokmål* and *riksmål*. Riksmål is a norm supported especially in linguistically conservative circles, and closer to the Danish written standard than bokmål, disallowing some forms spoken by many modern Norwegians. Bokmål is the official standard, broader in scope with respect to conservative and radical forms, radical meaning closer to spoken Eastern Norwegian. In the years following World War II the linguistic climate in Norway was at it hottest, especially between riksmål and bokmål, mainly due to a politically steered effort to establish a standard assimilating Standard Norwegian and Neo-Norwegian, called *samnorsk*, Common Norwegian.

Consequently, at its most complicated, written Norwegian has had four standards – nynorsk (which again includes several unofficial standards), bokmål, samnorsk and riksmål – representing separate ideologies and goals in language planning and also in their understanding of the political value of linguistic variants. Also minor standards have been launched, but they have had little influence on the modern standardization of written Norwegian and will not be taken into account here.

The BRO-initiative is meant to gather and unite the different standards that are not Neo Norwegian in one dictionary.

## 2. The fundament for building a linguistic bridge

In 2005, a revised written standard for bokmål was launched, reducing the difference between bokmål and riksmål to a minimum, mostly by incorporating the majority of the previous riksmål forms in the wider scoped bokmål.

Happily, it seems that the 2005-reform has led to a peace settlement of sorts for the majority variant, and paved the way to cooperation in documenting the vocabulary in one common dictionary project called BRO ('bridge', being an acronym for *Bokmålets og Riksmålets Ordbase*, meaning 'The vocabulary database of Standard Norwegian'). Hopefully, BRO will channel resources towards exploring and documenting the lexical inventory of modern Norwegian away from the old language disputes and political disagreements.

The BRO-project is a bridge resting on two piers. The one pier is the six volume *Norsk Riksmålsordbok* and its expanded and modernized new version *Det Norske Akademis Store Ordbok* (The Norwegian Academy's Dictionary, abbreviated *NAOB*). NAOB will be an Internet dictionary, is in full editorial progress, and aims – given sufficient funding – to be completed in 2014.

The other pier of the bridge is being provided by The Department of Linguistics and Literature, Section of Lexicography at the University of Oslo. It comprises scientifically based modern lexicographical tools like a balanced dictionary text corpus: *Leksikografisk bokmålskorpus* (Lexicographical Corpus of Standard Norwegian, abbreviated *LBK*), the morphological database *Norsk ordbank* (Norwegian Word Bank) and *Bokmålsordboka*, a comprehensive documentation of the modern Standard Norwegian norm with regard to morphological variant forms, since all forms in the official standard are given.

The formal BRO agreement was signed in August of 2011.

Supported by these two piers, the project hopefully will succeed in bridging the gap which exists in the lexicographical documentation of the modern majority standard of the Norwegian vocabulary.

### 3. The pier from the riksmål/NAOB side

NAOB is based on *Norsk Riksmålsordbok* (NRO), as of today Norway's only complete multi-volume dictionary, giving an abundance of literary quotes from the 19th and 20th centuries. The four main volumes of *Norsk Riksmålsordbok* were published between 1937 and 1957; two supplementary volumes in 1995. *Norsk Riksmålsordbok* contains a vast amount of valuable lexicographical information but requires extensive revision in order to function as a modern dictionary.

The revision effort includes language modernization (moderate *bokmål*), semantical modernization, simplification of complex hierarchies of meanings, adjustment of politically or socially obsolete viewpoints, pronunciation of non-compound lemmas, brief etymologies for root words, quotes from recent Norwegian literature, and addition of tens of thousands of entries and meanings.

More specifically:

- Language modernization: Implementing modern spelling forms throughout the vocabulary, e.g. eye: *øie* > *øye*, cloth: *klæde* > *klede*. Also implementing modern syntax throughout, most typically in the definitions.
- Simplification of meanings: The old NRO tends to define entries encyclopedically, giving far too many and too specific defining criteria. This results in many, unnecessarily narrow meanings and confusing, complex semantical structures. NAOB has less complex meaning structures.
- Addition of tens of thousands of entries and meanings: NAOB's aim is to cover modern Norwegian well enough to give the general public a dictionary which is perceived as up-to-date. As the funding scheme and BRO collaboration leave thorough corpus comparison to the ILN, NAOB makes use of substitute methods for covering new words and meanings: systematical gathering from sources such as encyclopedias and other dictionaries, as well as excerpting from modern literature.

NAOB's content is xml-coded, parsed by a stringently lexicographical DTD, resulting in articles with an easily understandable publishing interface. NAOB will offer full Internet functionality and a state-of-the-art technical platform for future updating.

NAOB intends also to involve the general public through contact on the web site, offering an interface for suggesting new content and improvements.

### 4. The pier from the bokmål/UiO side

Avdeling for bokmålsleksikografi (Section for Standard Norwegian Lexicography), University of Oslo is committed to documenting the complete official Standard Norwegian. This is mainly done in the one-volume dictionary *Bokmålsordboka*, where all allowed forms and variant forms are accounted for in full, as well as short and accurate definitions of all lemmas. As a spin off product of *Bokmålsordboka* and of some other resources, UiO also has developed *Norsk Ordbank*, which is a list of all lemmas with full flexion forms. Added are some words from IBM's lists and argument structure of verbs from NorKompLex, an early version of a computational lexicon for Norwegian.

The Section for lexicography at UiO not only compiles and edits dictionaries, but also focuses on general scholarly aspects of lexicography, and has for that purpose established a balanced, electronic text corpus aiming at 100 mill words, called *Leksikografisk*

*bokmålskorpus* (LBK) for lexical research. This corpus is already in use for compiling examples in NAOB, which is a first step in collaboration inside BRO.

To judge right or wrong in bokmål – good or bad, acceptable or unacceptable – BRO will be using the LBK-corpus<sup>1</sup> to find out which forms are in use, both in oral and written bokmål, as well as which forms are most used in different contexts and text types. A corpus based documentation of modern Standard Norwegian will thus be more objective and without any political bias, which is an important goal for the BRO-dictionary.

Having received fresh funding, the LBK-corpus is in good progress and will be completed in 2013 as a 100 mil. word balanced text corpus.

In addition to LBK, the Norwegian Speech Corpus can give some further information to the actual use of the variant forms. Some of the most common variant forms are also checked in the NoTa-corpus<sup>2</sup> for a short comparison, presented in table 2.

Fjeld (2008a) documents an extensive plan for a lexical description of modern bokmål language and the use of the resources developed at the University of Oslo. These plans will be incorporated in the BRO-project.

## 5. Bridging the gap

Bridging the gap between two piers with different history and ideology, presents challenges to the BRO parties. Future discussions and projects will help us determine how to best meet these challenges in practice.

In the Norwegian language, there are two fundamentally different ways of indicating your language political standpoint: a) by orthographic variants of the word forms, mostly diphthongs or monophthongs (e.g. *sein/sen*) or other deviations in the vowel system (dyp/djup) or in what derivation elements you chose (e.g. -ing/-ning); and b) through alternative flexion systems, like the two- or three-gender system for nouns, and past tense of verbs with *-a* or *-et*-derivation.

The first step needed to fulfill the requirement from the funding Ministry of Culture is to merge the lemma lists into one. Then through systematic corpus comparison using texts and tools available, we will fill in any lacunas found in the information types: lemmas, collocations, definitions and literary examples.

A way of presenting all-inclusiveness is to mark up variants in the xml code, enabling users to toggle between dictionary variants according to preference, using the style sheet. This marking could give the users options between different substandards, like so called moderate bokmål, or only the radical forms or so called rural forms. The interface could be constructed so as to show only the variant forms preferred by the user. This point needs further elaboration within the project.

The moderate bokmål substandard used by NAOB for main lemmas, morphology and editorial language, makes a deliberate choice for the user among alternative forms, indicating stylistically what is commonly preferred in formal written Norwegian.

The official standard documented in *Bokmålsordboka* implies all forms with full meaning, as well as the possibility to mix their morphological variant forms in derivations and compounds according to your own likes and stylistic requirements, and very few hints about stylistic value or usability.

It is the BRO project's goal to make all bokmål forms available in the BRO-database, accompanied with documentation of the frequency in use of the variant forms. The full morphology is to be fetched automatically from *Norsk Ordbank*, where all words with full-fledged forms are listed. The frequency in use will be found in LBK, when the project starts in full, hopefully consisting of 100 million words.

We will in the following present the first sketches for this collaboration, which also will involve the art of balancing on lines stretched in the span between the two existing piers.

### 5.1. Lemmatization of variant forms

NAOB chooses one variant – preferably the one most commonly used in written, formal contexts – as lemma sign form (labeled the *NAOB norm*), i.e. the entry word for the main article.

A good example of the differences between these two lemma lists is, ironically, the Norwegian word for bridge, **bro**, which has a parallel form **bru**. The two spelling forms have equal status in the official standard. In an all-inclusive bokmål dictionary without a recommended subnorm, they would be presented as variants of the same lemma in their rightful alphabetical place, not as two separate lemmas with a reference from one to the other, indicating that one form is preferable to the other. A possible way of doing this would be like this:

**bro** el. **bru** + full article  
**bru** el. **bro** + full article

There are quite many such variant forms like *bro/bru* in the Standard Norwegian. The following examples list lemmas with variant forms inside the official norm in the five first pages in a small spelling word list (Fjeld 2008b):

*adkomst/atkomst* and many other words with the prefix *ad-/at-*  
*agelaus/ageløs* and all other compounds with the adjective *løs/laus*  
*akselbred/akselbrei*, *akselbred/akselbrei* and all compounds with *aksel/aksle-* and other compounds with *bred/brei* as parts  
*aktersegle/akterseile* and all other compounds with *segle/seile*  
*aleine/alene*  
*allfarveg/allfarvei* and all compounds with *veg/vei*  
*allmenndannelse/allmenndanning*

The corpus gives the following frequencies of the use of the variants:

**Table 1.** The frequency of some orthographic variant forms in bokmål.

	<i>LBK</i>	<i>NoTa</i>
atkomst	4	0
adkomst	23	0
aleine	609	62
alene	11875	105
brei	107	39
bred	1870	68
laus	139	1 (compound 'tanlaus' = toothless)
løs	4436	71
segle	0	0
seile	581	16

Table 1 shows that the difference in oral use between *alene/aleine* (alone) is much smaller than in written texts. A larger corpus of oral bokmål is hopefully developed, to give a better verification. Monophthong forms like *segle* (to sail) seems not to be in use in bokmål and might be deleted from the norm. However, as long as the Language Council of Norway accepts a form, the BRO-dictionary will document it. The documentation we find, will of course be available for the standardization of the norms, and we hope for a closer cooperation with the Language Council giving the Council a more scientific basis for their decisions.

There are also some extra problems with derivations where both the first element word has orthographical variant forms and the derivation morpheme also has variant forms, like *fordyping/fordjuping/fordypning/fordjupning* (hollow, depression).

NAOB has lemmatized only the so called moderate forms, here just *fordypning/fordyping*. In the comprehensive BRO-dictionary the total universe of variant forms will have to be lemmatized in one common list, which in some instances will give a large number of sign forms of the same lemma. One of the more extreme examples is *høglønna* (highly paid), with its 12 variant forms:

**høglønnet**  
**høglønt**  
**høglønna**  
**høglønnet**  
**høglønt**  
**høylønna**  
**høylønnet**  
**høylønt**  
**høylønna**  
**høylønnet**  
**høylønnt**  
**høylønt**

The large number of variant forms is due to a mix of orthographical variant forms in both parts of the compound, and the different flexion possibilities. All of them are stylistically marked and will have to be documented in the BRO-database.

## 5.2. Noun Gender and Noun Morphology

Some parts of speech present more morphological deviation than others. In the nouns there are three genders in bokmål, masculinum, femininum and neutrum, and two in riksmål, common gender (= masculinum) and neutrum, like the example of *bro/bru* also showed above. This is one of the main differences between bokmål and riksmål, as riksmål sought to continue the tradition from Danish, while bokmål also seeks to account for the Norwegian dialects and the common used language system. Bokmålsordboka lemmatizes approximately 10700 feminine nouns, all needing such double gender marking and double morphological tables.

**bru** el. **bro** [f1](#) el. [m1](#)

f1 means femininum class 1, inflected according to this pattern:

brua	broa
bruer	broer
bruene	broene

m1 means masculinum class 1, inflected:

broen	bruen
broer	bruer
broene	bruene

NAOB avoids presenting the feminine inflection form for *bru*, in accordance with the general NAOB standard. In NAOB the feminine form is accepted in only a handful of nouns, like *jente*, *øy*, *bikkje*, which are seen as typically oral, or in some instances rural Norwegian words, never having been used in Danish.

The NoTa-corpus gives four search hits on the form *broa*, four hits on *brua*, no hits on *broen* and no hits on *bruen*. This corpus, unfortunately, is very small, 900 000 words only, so it just gives an indication of what is used most frequently in spoken modern Norwegian. LBK, and NoTa give the hits accounted for in table 2.

**Table 2.** Use of the different forms of bro/bru; broen/bruen/broa/brua in modern Norwegian.

	<b>bro</b>	<b>bru</b>	<b>broen</b>	<b>bruen</b>	<b>broa</b>	<b>brua</b>
<i>written</i>	696	207	900	2	98	371
<i>spoken</i>	1	2	0	0	4	4

We see that the NAOB-standard still covers the most frequently used forms, but according to modern oral language, this is in a rapid change. A dictionary like the BRO dictionary should account for such variants and be updated on a regular basis. In this way, the dictionary would also document the linguistic change in modern Norwegian.

### 5.3. Verb morphology

Bokmål allows alternative ways of flexing verbs, especially in the weak forms, while the NAOB standard chooses the flexion most commonly found in written bokmål:

NAOB:	<b>kaste</b> (to throw) - kastet- kastet
Bokmålsordboka:	<b>kaste</b> (to throw) - kastet- kastet kaste - kasta - kasta

The optional *-et* and *-a* in past tense and the participles, is the most striking difference between the two standards, and BRO has to give full morphology for all these verbs.

A great many of the weak participles are used as adjectival forms, as *bortkasta* (waisted) for bokmål and *bortkastet* for riksmål. According to concord flexion, the problem grows in pluralis, since the *a*-variant can be both singularis and pluralis, but the *et*-variant can be *-ete* or *-ede* in pluralis, in the official norm, only *-ede* in the NAOB-norm (cf 5.1 above):

**Table 3.** Use of the different participle forms in written and spoken Norwegian.

	LBK	NoTa
bortkasta	20	5
bortkastet	179	2
bortkastede	24 } 203	1 } 3
bortkastete	3	0

The investigation shows that the NAOB forms are the ones absolutely most used in written language, the other forms in spoken language.

#### 5.4 Sublemmas

Most lemmas will have superordinate lemmas or sublemmas. Under the lemma for **bro/bru**, Bokmålsordboka has three sublemmas under the concrete meaning:

gangb- (walkway), hengeb- (suspension bridge), vippeb- (bascule bridge)

and three under the metaphorical meaning,

kommandob- (command bridge), kuldeb- (cold thermal bridge), låveb- (barn ramp), varmeb- (warm thermal bridge)

NAOB has three others:

kjørebro (driving bridge?), gangbro (walkway), elgbro (moose bridge)

In the BRO-project we have to find out which of these sublemmas are the most important, The LBK-corpus has the most frequent compounds with *-bro*

kuldebro (84), hengebro (27), gangbro (23), jernbanebro (16), luftbro (11), trebro (8)

and with *-bru* as the second part of the compound

triangelbru (13), hengebru (7), gangbru (6), låvebru (4), trebru (wooden bridge) (4)

If the sublemmas are meant to show the most common compounds and hence the meaning of the lemma **bro/bru**, the corpus shows that the concrete meaning is best illustrated with **gangbro/-bru** to explain what a bridge is used for, and **hengebro/-bru** to show how a bridge often is constructed. The extended meaning is shown with the sublemmas **låvebro/-bru** and **kommandobro/-bru**. The metaphorical meaning is illustrated by the compounds **luftbro/-bru** and **kuldebro/-bru**. The different types of meaning should be ordered and sorted in the definitions like this

1. concrete meaning
  - purpose
  - material<sup>3</sup>
2. extended meaning
3. metaphorical meaning



Then the most frequently used sublemmas should be chosen to fill the different types. All sublemmas will normally have all variant forms documented.

### 5.5. Article structure and article size

The amount of examples and citations are very differently documented in the two dictionaries, and the terminology is not the same. This has historical reasons. A primary goal of the predecessor of NAOB, *Norsk Riksmålsordbok*, was to document the vocabulary as it was used by well-known and respected authors from the years after the independence from Denmark; it was perhaps intended as a literary dictionary more than a dictionary of the commonly used language. In NAOB both the literary and the common language is to be documented. Bokmålsordboka is in the first place meant to document the modern, common language. These resources sum up to be quite a good documentation of the Norwegian vocabulary as a whole, but in the first place, the terminology has to be adjusted and harmonized, and the different types of citations and language attestations have to be sorted in the same way, as well as multiword constructions and sublemmas. The ordering of the information types within an article should be:

1. Main lemma
- (2. Sublemmas)
3. Collocations
  - 3.1. Grammatical collocations
  - 3.2. Semantic collocations
4. Authentic citations/attestations
5. Idioms

The differences between the two dictionaries regarding multiword expressions can also be illustrated with the word **bro/bru**.

In Bokmålsordboka there are three multiword expressions listed for the concrete meaning of *bro/bru*:

1. *b-a over elva* (the bridge over the river)
2. *bygge b-* (build a bridge) også overf: få i stand tilnærming mellom to parter med helt ulike standpunkter
3. *bryte alle b-er (bak seg)* (break all bridges) ta en avgjørelse som ikke kan gjøres om igjen

Number one is a traditional example, showing the semantic meaning of the lemma. Number two is a semantic collocation, and number three is more an idiom than an example. They should therefore be marked as different information types. Another problem is to find the most frequent expressions. A corpus search shows that *bryte alle broer* has only one hit in LBK, but the idiom *brenne alle broer* (burn all bridges) has 19 hits and would have been more adequate in a dictionary documenting common use.

NAOB presents separate sections for collocations (called “examples”) and idioms (called “uttrykk”), potentially for every meaning.

In the BRO-project, a set of grammatical collocations will be documented for all relevant lemmas. They are to be found semiautomatically by means of corpus analysis of

LBK with the statistical program DeepDict Lexifier developed at GramTrans (Bick 2008). Such an analysis of the lemma *bro*, gives these results:

4.43:2 sprengje · 2.61:3 krysse · 4.5:1 vakte · 2.27:3 brenne · -0.85:6 bygge · 1.96:3 passere · 3.63:1 vedlikeholde · 2.72:1 reparere<sup>4</sup>

These figures mean that the most common V+N-collocation is *bygge bro* (build a bridge), the two next most frequent are *passere* (pass) and *krysse en bro*, which might be exchanged with ex. *legge* (lay), or be added in the BRO dictionary.

The statistical analysis with DeepDict shows that the most common collocations containing *bro*, are:

V+N-collocation: *bygge, passere, sprengje, krysse, brenne en bro*

A+N-collocation: *sukkenes bro, usynlig bro, provisorisk bro, smal bro, gigantisk bro*

N+PP postmodifiers: - *over kløft, elv, avgrunn, vollgrav...*

The lexicographic resources available for the project have to some extent overlapping multiword expressions, to some extent different ones documented. What seems clear, is that there is no good system of categorizing constructions, collocations, idioms and other multiword lemmas. This field needs refining according to a clearer system in the BRO-dictionary.

The small corpus explorations presented here, show that the forms and expressions documented in NAOB are generally more often used than the official ones in writing, but the official variant forms are used more often in spoken language. The BRO-dictionary will account for these differences for all lemmas, as a full-fledged corpus based dictionary.

## 6. Merging two different types of dictionaries into one common electronic version

There are still several questions to be answered before the project can be carried out. In this planning period, methodological issues will be explored and documented: How does one proceed in making a modern, extensive dictionary based on two existing and different ones? Different in scope, style and ideology. Is it possible to present a joint venture like BRO as a single-choice language norm based on the single user's feeling of national identity?

Another challenge is to find out how such an all-inclusive dictionary should be structured for the user in an Internet presentation.

Foreseeing several problems and obstacles, we are optimistic and think that this project will provide the documentation of the main variety of Norwegian vocabulary with its authoritative solution. The signing of the contract of cooperation between The Norwegian Academy of Language and Literature and The Department of Linguistics and Nordic Studies at the University of Oslo was signed 17th of August 2011, was an important event in our language history. We will do our best to live up to this.

## 7. BRO as a result of Norwegian language history and national identity

The wide norms for orthography and morphology in Norwegian are a heritage from the Danish linguist Rasmus Rask, who launched the so called orthophonic principle in orthography in 1826, which had great impact on Norwegian language planning, as we can see from the first Norwegian standardized norm from 1862 (Omdal og Vikør 1996). Since

Norway is a large, rugged and sparsely populated country, it has developed more dialect variation than most other nations. To unify as many dialects as possible inside one standard, the first norm from 1907 was made relatively open for orthographic variant forms.

The dialects or spoken written standards also came in use among the establishment in the central cities and through the 20th century, identified as upper class sociolects. Which forms you chose inside the open norm showed which social class you belonged to. Consequently, determining which forms were accepted inside the norm, was not only a linguistic issue, it was also a general political issue that made way for the famous – and to other nations mystifying – Norwegian language struggle. Norway is a community with small differences in wealth and status, and we have had much less of an aristocracy to set its indisputable upper class sociolect as a standard norm for orthography and phonology, than most other European linguistic societies. In Norwegian, all dialects are generally accepted in all ways of life, and pronunciation is seldom given in dictionaries, to avoid any discrimination or elitism. Many poets use their own dialect in their lyrics, arguing that their dialects are their "heart language" enabling them to express their inner, personal voice and attitudes in the best way.

To edit a common dictionary presenting all norms and subnorms in such a linguistic climate is a difficult task. In light of this we hope to make an interface for the dictionary where the users may make their own choices between several substandards before entering the dictionary. This, of course, must not prevent the interface from showing all variants, among other things a very useful tool for teachers when correcting their pupils' work.

It is said that Norway is a society of individualists where all members are equal. This seems like an utopia, but the BRO-dictionary will try to satisfy all lexicographical requirements to embody this typically Norwegian form of national identity.

## Notes

<sup>1</sup> [http://www.tekstlab.uio.no/glossa/html/index\\_dev.php?corpus=bokmal](http://www.tekstlab.uio.no/glossa/html/index_dev.php?corpus=bokmal)

<sup>2</sup> <http://www.hf.uio.no/iln/tjenester/kunnskap/sprak/korpus/talesprakskorpus/nota-oslo/>

<sup>3</sup> Such sense division of nouns is inspired by Pustejovskis qualia structure

<sup>4</sup> The figures are from the DeepDict analysis. The two figures before the comma shows the relative binding between the two words in the collocation, relative according to their frequency in the total corpus and how often they appear together in this collocation. The figure after colon states the logarithmic frequency of the collocation in the corpus in total. For further reading of this analysis, cf. Bick 2009.

## References

- Bick, E. 2009.** 'DeepDict - A corpus based Dictionary of Word Relations'. In *Proceedings of Nodalida 2009*. Odense, Denmark: NEALT proceedings series, Vol. 4, 268–271.
- Fjeld, R. V. 2008a.** 'Plan for leksikalsk dokumentasjon av moderne bokmål.' In Á. Svavrsdottir et al. (eds.), *Nordiske Studier i Leksikografi 9. Rapport fra konference om leksikografi i Norden, Akureyri 22.-26- mai 2007*. Reykjavík.
- Fjeld, R. V. 2008b.** *Lingua Bokmålordliste med skrivereregler*. Oslo: Kunnskapsforlaget.
- Pustejovsky, J. 1995.** *The generative lexicon*. Cambridge, Mass.: MIT Press.
- Omdal, H. & L. S. Vikør 1996.** *Språknormer i Norge. Normeringsproblematikk i bokmål og nynorsk*. Oslo: Cappelen Akademisk.