

COMMUNICATION & LANGUAGE at work

Issue no. 1 | Summer 2012

ICT Tools and Professional Language

Using Web-Based Corpora to Find Norwegian Specialised Neologies

Marita Kristiansen

(pp. 11 – 20)

<http://ojs.statsbiblioteket.dk/index.php/claw/article/view/7235>

Subscribe:

<http://ojs.statsbiblioteket.dk/index.php/claw/notification/subscribeMailList>

Archives:

<http://ojs.statsbiblioteket.dk/index.php/claw/issue/archive>

Publishing:

<http://ojs.statsbiblioteket.dk/index.php/claw/about/submissions#onlineSubmissions>

Contact:

<http://ojs.statsbiblioteket.dk/index.php/claw/about/contact>



| Bridging Theory and Practice |

<http://ojs.statsbiblioteket.dk/index.php/claw>

Using Web-Based Corpora to Find Norwegian Specialised Neologies

Marita Kristiansen

Associate Professor

Department of Professional and Intercultural Communication

NHH – Norwegian School of Economics

Abstract

In this article I present an on-going study of the use of web-based corpora to detect Norwegian terminology and in particular financial neologisms. With an ever increasing influence from English in the domain, the use of anglicisms such as 'dobbeldipp' ('double dip') or 'shortsalg' ('short selling') is increasing rapidly. The study focuses on how to detect these specialised neologisms and also possible Norwegian terms that are formed. Three different corpora are used, i.e., the Norwegian Newspaper Corpus (NNC), Peter Warrens finansblogg (PW) and a blog on a research project which focuses on the financial crisis called Crisis, restructuring and Growth (KOV). The aim is to gather terminological information which can be useful in the NHH Termbase, an online, web-based termbase to be used by students, lecturers and researchers at the Norwegian School of Economics (NHH).

1. Introduction

The language used for mediating specialised communication is to an increasingly extent English. Following the Bologna declaration of 1999 which promotes a European area for higher education with focus on easy transfer of students, lecturers and researchers across countries, the pressure from English has increased even more. This raises the question of whether we are heading towards domain loss in some disciplines. In Norway this concern is expressed for instance in a comprehensive report (Norsk i hundre) prepared by the Norwegian Language Council, which discusses possible strategies for how to maintain Norwegian as the preferred language, and in a white paper on national language policies (Report no. 35 (2007-2008) to the Storting). In line with these two reports, a recent amendment to the Norwegian act relating to universities and university colleges gives the responsibility to maintain Norwegian terminology to the institutions (Ot.prp. nr. 71 (2008-2009)). Domain loss in this connection is understood as the situation in which another language (in this case English) is becoming the preferred language of mediation in a given domain and through this, forcing the native language to decline or even disappear.

In this article I will discuss methods for detecting Norwegian specialised neologisms in popularised texts, including newspapers and domain-related blogs which are readily available on the web. The purpose is to find Norwegian terminology which may not be

included in for instance textbooks yet. To illustrate the methodology I have applied, I will use finance as an example, with focus on concepts which have been particularly salient in connection with the recent financial crisis.

By specialised neologisms I here refer to terminology and expressions which are used in connection with finance and not new words in general. Also, the analysis has a particular focus on anglicisms, that is, expressions which originate from English, but which have been adopted in Norwegian texts. The analysis thus focuses on English loan words (anglicisms), such as ‘subprime’ or ‘short selling’, calques, including among others ‘shortsalg’ (‘short selling’), ‘dobbeldipp’ (‘double dip’), and ‘CDO-obligasjon’ (‘CDO bond’), or new Norwegian terms, such as ‘sikringskontrakt’ (‘hedge fund’) or ‘kredittskvis’ (‘credit crunch’).

2. Why searching the web for new terms

In recent years I have been working on developing an internet-available termbase for students and lecturers of economic-administrative domains (NHH Termbase, Kristiansen 2010). To develop this resource I have based the data extraction mainly on textbooks aimed at intermediary students to capture the most central concepts and terms in the various disciplines. A motivation behind the project has been to make Norwegian and English terminology, including concept structures and definitions, readily available to the students. The goal has been to provide the students with a learning tool, by bridging the two languages and linking the terminology to central concepts and definitions. Also, such a tool will establish a parallel terminology in line with the responsibility given to NHH to establish economic-administrative terminology through the legal amendment mentioned above.

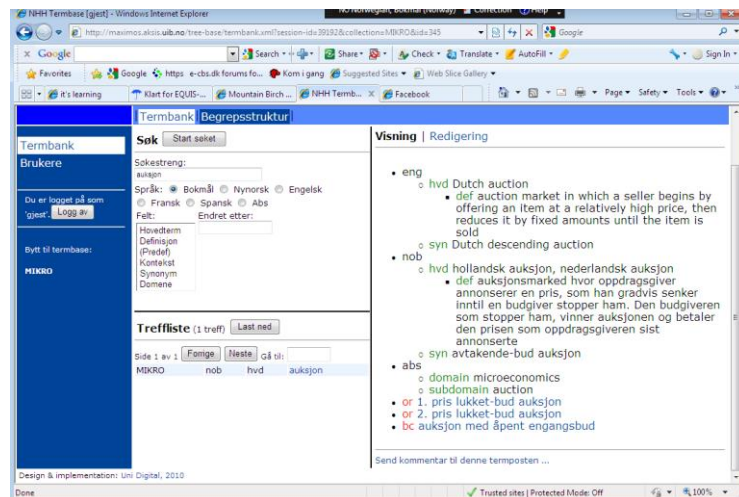


Figure 1. NHH Termbase – prototype interface

However, updated textbooks are not always available, for instance when domains or subdomains are characterised by rapid conceptual and terminological changes. An example may be the discipline *financial accounting*. Financial accounting has been

subject to international harmonisation through the International accounting standards (IAS) and the International financial reporting standards (IFRS) adopted by the EU in 2002. This has resulted in both concept and terminology changes, for example a shift towards focusing on the concept of *fair value* and changing the term ‘profit and loss account’ to ‘income statement’. Another discipline is *finance*, which is characterised by rapid market changes, the development of new financial products, such as structured savings products, and is also subject to international standardisation to achieve cross-national harmonisation and transparency (Kristiansen 2011; Kristiansen and Andersen in press). Assuming that financial specialists and researchers as well as the media have wanted to discuss the financial crisis in recent years, I have investigated three different corpora, namely the already existing *Norwegian Newspaper Corpus* (NNC),¹ *Peter Warrens finansblogg* (PW; a finance blog)² and NHH’s researchers’ blog on *Crisis, restructuring and Growth* (KOV).³ The two latter corpora have been established for the purpose of this study in particular.

In the following I will present the methodology and these three corpora which have been applied to harvest Norwegian financial neologisms related to the financial crisis. Next, I will give some examples of the information that can be found in the corpora and discuss how this information can be useful as input into termbases, such as the NHH Termbase.

3. The Norwegian newspaper corpus (NNC) as a starting point

The methodological point of departure has been the NNC. The NNC, established in 1998, is a self-expanding corpus consisting of Norwegian newspaper texts (close to 900 million words). It is updated automatically on a daily basis and comprises 10 of the most central newspapers in Norway. In my analysis I have used the NNC to search for term variants used for already known concepts such as *subprime* or *financial crisis*. For instance, the analysis shows that as many as 24 term variants (also including orthographic variants such as ‘høyrisiko boliglån’/‘høyrisiko-boliglån’ (‘high-risk mortgage’)) are used to denote the concept of *subprime lending* in the NNC. I will return to this concept in the discussion of findings in section 3. In addition to providing term variants, the NNC also offers information which can be used to establish both definitions and concept relations as illustrated in [1]-[3].⁴

- [1] Børsnoterte fond, også kalt Exchange Traded Funds (ETF) er aksjefond som er konstruert for å følge avkastningen til en aksjeindeks, ... (VG101110)
Exchange traded funds (ETF) are unit trusts which are constructed to track share index prices

¹ <http://avis.uib.no/>

² <http://www.peterwarren.no/>

³ <http://blogg.nhh.no/kriseprogrammet/>

⁴ In the examples, the names of the newspapers are indicated together with the date of publication, e.g. “AP100916”, AP here referring to the Norwegian daily *Aftenposten*. For an overview of which newspapers the initialisms refer to, see <http://avis.uib.no>.

- [2] En ETF er et fond som består av aksjer og andre verdipapirer... (AP100916)
An ETF is a fund consisting of shares and other securities...
- [3] ... børshandlede aksjefond, som også kalles ETF - Exchange Traded Fund... (AP050919)
... exchange traded funds, which are also called ETFs ...

From these examples we get the English term ('Exchange Traded Fund'), its acronym ('ETF') which, based on the NNC, is used frequently and consistently in Norwegian as well. In addition, we get two competing Norwegian equivalents for 'exchange traded fund', i.e., 'børsnotert fond' and 'børshandlet (aksje)fond', respectively. Furthermore, information which can be used to develop a definition is also extracted: "konstruert for å følge avkastningen til en aksjeindeks" (constructed to track share index prices) and "fond som består av aksjer og andre verdipapirer" (a fund consisting of shares and other securities). A superordinate concept is also provided '(aksje)fond' ('unit trust'). See Kristiansen 2012, 2011 for more details on relational information in the NNC.

In addition to the search interface used to find lemmas within the range of newspapers in the NNC, a neology detector has been developed which harvests all new word forms used daily based on a comparison with already registered forms (some 5.3 million word forms). The comparison is also made towards a full-form lexicon derived from the comprehensive dictionary of Norwegian bokmål, *Bokmålsordboka* (Andersen and Hofland 2012).

4. Blogs – a repository of neologisms

In order to find specialised neologisms this detector has been applied on two other corpora. The first corpus, *Peter Warrens finansblogg* (PW), is a finance blog written by a subject specialist and covers current topics related to, for instance, hedging, markets, commodities and relevant macroeconomic topics. Thus since this blog is in Norwegian and is focusing on topics which are highly global and typically discussed in English also by Norwegian specialists I believe it will be a good source of Norwegian terminology. The blog was established in March 2009 and by 25 February 2012 as many as 625 neological forms have been detected.

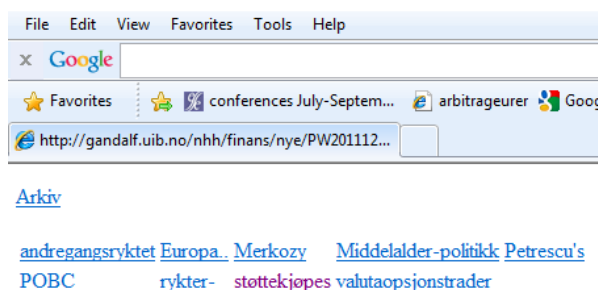


Figure 2. Neologies detected from PW on 10 December 2011

Examples of relevant specialised neologisms that have been detected are English loan words in Norwegian such as ‘Non-Deliverable Forward’, ‘asset’ene’ (‘the ... assets’), calques such as ‘tradingprogram’ (‘trading programme’) or ‘spreadmessig’ (‘regarding spreads’), or Norwegian neologisms such as ‘ultrakortsiktige’ (‘ultra short-term’) or ‘nasjonalgjeldskrise’ (‘national debt crisis’). The latter relates to the on-going economic crisis in Europe and is closely related to that of ‘statsgjeld’ (‘sovereign debt’).

- [4] Finanskrisen etterfulgt av en **nasjonalgjeldskrise** i Europa har gjort at den etterlengtede oppgangen i den vestlige økonomien har latt vente på seg. (PW2010/09)
The financial crisis followed by a **national debt crisis** in Europe has led to a delay in the much longed for economic upturn in the Western world.

However, also general neologisms are detected, such as new forms; ‘brannforsikre’ (‘to insure against fire’; verb form), as well as spelling errors ‘allo-ker-ingen’ (‘the allocation’), abbreviations ‘OTC-produkter’ (‘OTC products’), name candidates (‘SPGSCI’, ‘Goldman Sachs-kopi’ (‘Goldman Sachs copy’), ‘Merkozy’, cf. [5], and also some cases of garbage, e.g. noe” (some”).

- [5] For **Merkozy** er slike kostnader for trivielle å regne dersom det sikrer dem gjenvalg. (PW 2011/12)
For **Merkozy** such expenses are trivial if it ensures their re-election.

The second corpus is a researchers’ blog developed at NHH with blog texts concerning a cross-disciplinary research project called *Krise omstilling og vekst* (KOV; *Crisis, restructuring and growth*). The project was established in 2009 and runs for five years. It focuses on the causes of the recent financial crisis, its impact on future restructuring needs and prospects for growth. Key areas are the financial crisis, crises in general, macroeconomic perspectives, restructuring, financial stability, economic policy, economic growth and structural innovation. The NHH researchers will normally discuss their findings in scientific presentations and articles using English. However, in this blog they are forced to use Norwegian when they explain their studies. Consequently, this text source should also provide interesting terminological information which could be relevant for the termbase.

Neologisms detected in the KOV blog include English loan words such as ‘copulas’ ([6] ‘copulas’), calques, such as ‘Thanksgiving-salg’ ([7] ‘Thanksgiving sale’), or Norwegian “substitute words” like ‘bankrelasjon’ (‘bank relation’) or ‘finanskrisemålene’ (financial crisis targets’). See [15]-[16] in section 3 for examples of the two latter expressions.

- [6] En metode som er meget populær for å modellere smitteeffekter i finansielle markeder er såkalte **copulas** (KOV230211).
A method which is very popular to model contagion effects in financial markets is the so-called **copulas**.

- [7] For å sikre seg en enda lavere pris ventet han til det var **Thanksgiving-salg** i butikkene (KOV050111).
To get an even lower price he waited until the **Thanksgiving sale** in the shops...

Many of the expressions in the KOV corpus denote concepts which are more theoretical, such as ‘copula’ (cf. [6]), and which are not likely to be used in a general language setting. In teaching these expressions are, nevertheless, central. The expression ‘copula’ is already established in the Norwegian language, however, spelled ‘kopula’. Originating from *linguistics* it refers to a word which is used to link the subject of a sentence with a predicate. In KOV it is used in connection with analyses of financial markets. The use of an anglicism in this connection indicates a shift from Norwegian to English.

5. Term variation, neologisms and relational information

In the following I will present some results from the analysis which shows that both the newspaper corpus and the blogs detect many neologisms which are relevant in developing a Norwegian specialist language in finance, beginning with neologisms related to the concept of *subprime*.

5.1 The use of ‘subprime’ in Norwegian newspapers

The concept of *subprime* is perhaps the one that was most discussed in the early stages of the financial crisis. The anglicism ‘subprime’ first appeared in the NNC in March 2007. Since then more than 24 terms or explanatory paraphrases have been used as variants of ‘subprime’. In examples [8]-[10], three different synonyms are used to express the concept of *subprime lending* in Norwegian. What should be noted is that the relatively colloquial ‘råtne boliglån’ (‘rotten mortgage’) [9] is the only expression which is consistently used alone without the addition of ‘subprime’, in contrast to what is the case with for instance ‘høyrisikolån’ (‘high-risk loan’) [10] (Kristiansen 2012).

- [8] Banker ble pålagt å gi såkalte **subprimelån** til låntagere uten egenkapital. (AP081119)
Banks were forced to offer so-called **subprime loans** to borrowers with no equity funding.
- [9] **Råtne boliglån** kan skremme flere investorer fra aksjemarkedet (DN070305)
Rotten mortgages may scare several investors off the stock market
- [10] I året som fulgte, begynte utlånere av **høyrisikolån** (såkalte "**subprimelån**") å melde seg konkurs. (DA090106)
In the following year, lenders of **high-risk loans** (so-called **subprime loans**) declared themselves bankrupt.

By counting the frequency of each variant per year, it is evident that some expressions are more used than others, as illustrated in Figure 3. Figure 3 displays five different variants; ‘subprimelån’ (‘subprime loan’), ‘subprime boliglån’ (‘subprime mortgage’), ‘høyrisiko-boliglån’/‘høyrisiko boliglån’ (‘high-risk mortgage’), ‘råtne boliglån’

(‘rotten mortgages’) and ‘lavt gradert lån (subprime)’ (‘low-graded mortgage’). As the graph shows, ‘subprimelån’ and ‘råtne boliglån’ are by far the most frequent, and these two variants compete in frequency from year to year. The latter is obviously more informal and carries a clear negative connotation not necessarily suitable in a formal specialised setting like for instance teaching. Also, it is clear that the calque ‘subprimelån’ is preferred compared with the Norwegian expression ‘høyrisiko-boliglån’, although the latter is more transparent. Overviews of usage, such as this one, should be used when deciding which of the terms should be recommended in Norwegian.

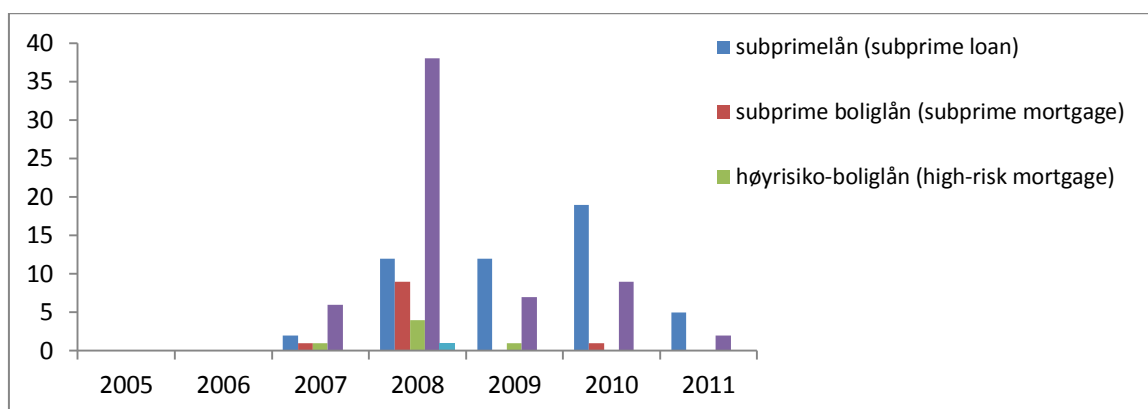


Figure 3. Term variants for the concept of subprime lending in the NNC

Towards the end of 2009 it looked like the concept of *subprime* would vanish, both from the NNC and from the global financial scene. However, as the financial crisis has unfortunately developed into an economic crisis, newspapers have started discussing the concept again, e.g. in connection with the concept of *sovereign debt*.

[11] I USA satt bankene på **råtten subprime**, i Europa sitter bankene på statsgjeld fra Piigs-landene (DN110808)

In the USA the banks took on **rotten subprimes**, in Europe the banks have taken on sovereign debt from the Piigs countries.

Whereas the concept of *subprime* may not defend an entry in a general language dictionary over time (and may not have a high enough frequency to be included), the relevance of the concept in a termbase, on the other hand, is evident from an economic perspective. In addition to the already mentioned expressions, other expressions have been formed, such as ‘subprimemarked’ (‘subprime market’) and ‘suprimeboble’ (‘subprime bubble’).

5.2 Specialist blogs ahead of the newspapers

Since I, as a terminologist, do not work actively within finance on a daily basis, it is difficult to track topical concepts and expressions regularly. The use of the neology

detector on the PW corpus, therefore, enables me to get an immediate overview of several topical concepts from which I may work to expand the NHH Termbase. In addition, the PW blog is often ahead of the more general NNC, as regards the introduction of neologisms in the domain. An example may be the expression ‘stressindeks’ (‘stress index’), which is related to that of ‘stresstest’ (‘stress test’). The concept of *stress test* describes how vulnerable the financial system is to different kinds of financial distress or shocks.⁵ The expression ‘stressindeks’ first emerged in the PW blog in June 2010, with reference to the EU index.

- [12] Vår egen EU **stressindeks** viser dette tydelig. (PW2011/4)
Our own EU **stress index** shows this clearly.

Examples of actual indexes may be the IMF financial stress index, which captures the most important episodes of financial stress in a set of emerging economies, or the EU financial stress index. The expression has now also emerged in the NNC (7 February 2012).

- [13] Norges Banks **stressindeks** er utviklet med utgangspunkt i den Hanschel og Monnin konstruerte for sveitsisk banksektor i 2005... (DN070212)
The **stress index** developed by Norges Bank (the Norwegian central bank) is based on the one developed by Hanschel and Monnin for the Swiss banking industry in 2005...

Another interesting neologism which has been detected in PW is that of ‘algofobia’ (‘algophobia’, i.e. the fear of pain) introduced in the PW as a label for the dislike in the innovative use of algorithmic or automated trading, or ‘sentralbankintervensjon’ (‘central bank intervention’, i.e., the buying or selling of currency by central banks).

5.3 Neologisms in researchers’ blogs relevant for teaching

As mentioned above, the neologisms detected in KOV are slightly more theoretical. A concept which is closely related to that of *credit crunch* and the fall of banks in the early stage of the financial crisis is that of *banking relations*. This concept is central in the KOV project and its research on how important banking relations are for enterprises’ access to credit, or lack of such.

- [14] Ved hjelp av detaljerte data om norske bedrifters banklån studerer vi derfor hvor viktige **bankrelasjoner** faktisk er for selskapenes tilgang til kreditt (KOV320511).
Based on detailed data on Norwegian enterprises’ bank loans, we therefore study how important **banking relations** in fact are for the enterprises’ access to credit.

⁵ <http://tinyurl.com/7s5flmx>

Another neologism which has been detected in the KOV blog is that of ‘finanskrisemål’ (‘financial crisis target’):

- [15] Vurderer vi dem 1 år før oppkjøpet ser vi dog at **finanskrisemålene** er bedrifter som er spesielt hardt truffet av finanskrisen ... (KOV250811).
When considering them 1 year before the acquisition, we see, however, that the **financial crisis targets** are enterprises which are particularly hard hit by the financial crisis...

Both of these expressions refer to concepts which are relevant in courses on financial crisis topics. As such courses will typically be offered in English since they are attractive to NHH’s foreign exchange students, it is relevant to present them with the Norwegian terminology through the NHH Termbase. The latter expression will, based on other neologies detected from the KOV blog, also be related to a set of others, such as ‘finanskrisetransaksjon’ (‘financial crisis transaction’), ‘kapitalallokering’ (‘capital allocation’), ‘konjunktursjokk’ (‘cyclical shock’) and ‘brannsalgshypotese’ (‘fire sale hypothesis’), all of which are related to the concept of *market for company control*. Together these expressions thus make up a subdomain of finance and can be registered as such in the termbase.

6. Concluding remarks

In this article I have discussed the use of web-based corpora to detect Norwegian terminology and in particular financial neologisms. As the influence from English increases in many domains there is a fear that Norwegian terms will no longer be developed, something which may result in domain loss. Thus three different corpora, including newspapers, a financial specialist blog and a researchers’ blog, have been used to detect specialised neologisms and to see whether Norwegian terminology can be found in addition to their English counterparts.

The analysis shows that these corpora provide much information which is interesting to register in the NHH Termbase. In addition to including many anglicisms, the NNC provides many calques and Norwegian equivalents. However, Norwegian terms are used alongside the English ones and hybrid terms continue in use throughout the period of analysis. This calls for an active language, or terminology planning in addition to the mere registration of data. As for the finance blog, PW, and the researchers’ blog, KOV, these also provide many relevant financial neologisms, which are often detected at an earlier date than in the NNC. Consequently, using these web-based sources seems to be a fruitful addition to the traditional textbooks.

As for the more overall question of domain loss, the findings from my analysis demonstrate that Norwegian terms do exist. However, what is lacking is a systematic and regular maintenance and dissemination of relevant terms at a national level.

Author



Marita Kristiansen

Associate Professor in English at NHH

Marita Kristiansen is associate professor in English at NHH (Norwegian School of Economics), where she teaches professional English and terminology. Her research interests include domain and term dynamics, neology, economic-administrative terminology and financial jargon. She has published papers internationally in e.g., Terminology, Studies in Corpus Linguistics and Terminology Science and Research.

7. References

- Andersen, Gisle and Knut Hofland. 2012. Building a large corpus based on newspapers from the web. In Andersen, Gisle. (ed). *Exploring Newspaper Language - Using the web to create and investigate a large corpus of modern Norwegian*. Amsterdam/New York: John Benjamins, 1-28.
- Crisis, restructuring and growth*. URL: <http://blogg.nhh.no/kriseprogrammet/>, 13.02.2012.
- IFRS – International Financial Reporting Standards (IFRSsTM) including International Accounting Standards (IASsTM) and Interpretations as at 31 March 2004*. London: International Accounting Standards Board.
- Kristiansen, Marita. 2012. Financial jargon in a general newspaper corpus. In Andersen, Gisle (ed). *Exploring Newspaper Language - Using the web to create and investigate a large corpus of modern Norwegian*. Amsterdam/New York: John Benjamins, 257-283.
- Kristiansen, Marita. 2011. Domain dynamics in scholarly areas. How external pressure may cause concept and term changes. In Temmerman, Rita and Marc Van Campenhout (eds). *The Dynamics of Terms in Specialized Communication. An Interdisciplinary Perspective. Terminology 17(1)2011*, 30-48.
- Kristiansen, Marita. 2010. Language Planning in Higher Education. The Case of Microeconomics. In Heine, Carmen and Jan Engberg (eds). *Online proceedings from the XVII European LSP Symposium 2009*. URL: <http://www.asb.dk/fileadmin/www.asb.dk/isek/kristiansen.pdf>. 1-13.
- Kristiansen, Marita and Gisle Andersen. In press. Corpus approaches to neology and their relevance for dynamic domains. To appear in *Neologica* 6/2012.
- Norsk i Hundre*. URL: http://www.sprakrad.no/upload/9832/norsk_i_hundre.pdf
- Norwegian newspaper corpus*. URL: <http://avis.uib.no>
- Ot.prp. nr. 71 (2008-2009) Om lov om endringer i lov 1. april 2005 nr. 15 om universiteter og høyskoler*. URL: <http://tinyurl.com/6wo2bj4>
- Peter Warrens finansblogg ((PW)*. URL: <http://www.peterwarren.no/>
- Report to the Storting no. 35 (2007-2008) Mål og mening. Ein heilskapleg norsk språkpolitikk*. URL: <http://tinyurl.com/79ng68t>
- Report to the Storting no. 13 (2005-2006) Kredittmeldinga 2005*. URL: <http://tinyurl.com/7s5flmx>