Towards automatic tracking of lexical change: linking historical lexical resources

Malin Ahlberg, Peter Andersson

Språkbanken / Department of Swedish
University of Gothenburg

malin.ahlberg@gu.se, peter.andersson@gu.se

ABSTRACT

In the field of historical linguistics, large-scale corpora studies are a key component in identifying phenomena such as language variation and language change. Manually performed corpora studies are very time consuming and may obscure interesting changes in the sense that phenomena that are not being specifically searched for easily are overlooked. During the last couple of years the potential of language technology tools has been put forward in relation to historical linguistic research. This paper is based on an experiment of linking up several lexical resources in Swedish, which together reflect a vocabulary from Old Swedish to Contemporary Swedish. The link-up aims at identifying potential lexical change such as cases of grammaticalization and it may further be of use in other language technology applications. In our case study we are linking lemmas together with part-of-speech information given in each entry for all the lexical resources. This paper describes our first results, where we focus on the cases when information about word class differs in one of the resources. In future studies it is necessary and desirable to include more digitalized lexicon resources and confirm the analyses with corpora research. Still, the current result already shows some interesting cases of semantic change and grammaticalization. Changes in the content word system such as generalization and specialization of meaning are also exemplified in our data. Even though the links sometimes show errors that at first sight lead us towards a wrong conclusion we believe that methods like the one used here may be very fruitful to future research to reach more efficiency in historical linguistics research.

KEYWORDS: lexical linking, historical linguistics, language change.

1 Introduction

During the last couple of years there has been an increased interest in the potential of language technology tools for other disciplines than modern linguistic research. Recent work (see e.g. Pumfrey and Mariani, 2012) has shown that large-scale corpus-based historical linguistics is a fruitful approach. An obvious area of collaboration between historical linguistics and language technology research is language change, such as grammaticalization. The work described in this paper aims at identifying some aspects of lexical change starting from an experiment on a link-up of a large bank of lexical resources.

The purpose of the article is twofold: first, we describe a link-up of historical lexica to modern resources, where existing lexica is integrated into a resource functioning as a dictionary between Old and Contemporary Swedish, suitable for language technology applications. It may be used to increase the readability of old texts, in information retrieval for historical corpora and for annotating and analysing these. Second, we give a case study of how large-scale comparisons of lexica from different time periods may help identifying and conceptualizing language change. Our intention is to give an example of how more efficient methods, in this case a set of automatically created links between historical and modern lexical units, can be used to support the formation of new research questions. Even if this approach generates some links without any apparent relation and obscures or misses others, the method may identify lexical change that we never have discovered with manual work. The case study in this paper examines the lemmas that exist in both historical and contemporary lexica and compares them with respect to which word classes they are assigned by different dictionaries.

Our central resources are five lexical resources available at Språkbanken[1]: Söderwall (1884), its supplement(Söderwall, 1953) and Schlyter (1887) for Old Swedish, Dalin (1855) for Modern Swedish, Saldo(Borin et al., 2008) for Contemporary Swedish. The rough distinction between language stages is based on the frequently used subdivision summarized in (Andersson, 2007). Söderwall's and Schlyter's dictionaries reflect Old Swedish (henceforth OSwedish) up to about year 1526 with the bible translation of The New Testament as a landmark. The period thereafter and up to about year 1900 is usually subsumed as Modern Swedish, for younger language the term Contemporary Swedish is used. We refer to Modern and Contemporary Swedish as MCSwedish.

The historical lexica have been digitized and are all available in LMF[2] format. They contain totally 50 000 entries of OSwedish, but the lexica overlap and the actual number of separate lexical entries is assumed to be lower. On the other hand, most entries contain a number of different senses, among which the semantic difference may be quite big. Saldo, a semantic and morphological lexicon of Contemporary Swedish, contains 120 000 entries. It is the central component in the Swedish FrameNet++ (SweFN++) (Borin et al., 2010), in which a number of lexical resources are integrated. Our work connects the three historic dictionaries to SweFN++ by a simple matching procedure presented in Section 3. In Section 4 we give examples and discuss cases of semantic change and grammaticalization that are shown when comparing the results of the link-up.

---

[1]http://spraakbanken.gu.se
[2]http://www.lexicalmarkupframework.org

2 Background

The examples of link-ups and integrations between modern lexical resources are many. Babelnet (Navigli and Ponzetto, 2012) is a multilingual ontology, constructed from an automatic integration of Wikipedia and WordNet. de Melo and Weikum (2009) combine a number of wordnets with information from dictionaries and parallel corpora into the Universal WordNet. Another example is Uby (Gurevych et al., 2012) combining English and German lexical resources.

For historical lexica, the digital resources are more scarce. Gotscharek et al. (2011) present an interactive tool for constructing lexicons from corpora, which they use to produce a historical lexicon for German with 10 000 entries. In information retrieval for historical corpora, most research however normalizes the text to modern spelling, Koolen et al. (2006) present some different methods for this. Ernst-Gerlach and Fuhr (2007) on the other hand transform a given search query in modern language into a number of possible historical forms. Borin and Forsberg (2011) present a diachronic lexical resource, where Dalin is linked to Saldo. We will refer to this material as the Diapivot. Their work further includes the start on a link-up between the OSwedish lexica and Saldo, which constitutes a starting point of our work.

In the case study of this paper we mainly focus on identifying semantic changes and grammaticaliziation. Semantic change includes processes such as generalization and specialization hence broadening and narrowing of meaning. Grammaticalization is the well described unidirectional process of correlating semantic, morpho-syntactic and phonological developments, often resulting in change of word class (Hopper and Traugott, 2003; Andersson, 2007). Traugott (2001) defines grammaticalization as "the change whereby lexical items and constructions come in certain linguistic contexts to serve grammatical functions or grammatical items develop new grammatical functions." Frequently cited examples are the development of auxiliaries from main verbs such as the change of the Swedish main verb *verka* 'act, work' to an auxiliary *verka* 'seem' or the adjective *bar* 'naked' which has developed to an adverb *bara* 'only' and further into a speech particle *ba* (Eriksson, 1992).

3 A (transitive) link-up from Old to Contemporary Swedish

The goal of the link-up discussed in this paper is to match the lemmas given in three OSwedish lexica to corresponding lemmas in modern lexica. The language has undergone much orthographical change and standardizations leading to a vast differences between Old and Contemporary Swedish. In order to find links between lemmas between which there is no longer a (visible) linguistic connection, we base the linking on information given within the entries. These often include a translation to Late Modern Swedish (see Figure 1), which we extract

äptirfinna vb	*återfinna.* "haffuer ther til badhä hug och sinne thz skiutast iach ma idher effter finnä" Iv 1912 .
värdhskylda vb	*förtjena.* "hwat ey wärdhzskylladhe war syndh wphöghilse, sannelica ney Skrifter till uppbygg. 73."
æpli nn	*äple.* VG. lIII. 123; U. Kk. l7: 5. not. 45; ME. * St. Eds. 25; Thj. 6.

Figure 1: Examples of lexical entries in the Old Swedish lexica

together with the lemma it self and its part-of-speech tag. The definitions are then compared to the lemmas in Dalin - if a match which in turn is linked to Saldo via the Diapivot is found, we create a link from the OSwedish lemma to the contemporary entry. Examples are shown in Figure 2a-b. The results are promising, we get 16 000 links of surprisingly good quality

	OSwedish		definition		Dalin		Saldo
a.	värdhskylda (vb)	→	förtjena	→	förtjena (vb)	→	förtjäna (vb)
b.	æpli (nn)	→	äple	→	äple (nn)	→	äpple (nn)
c.	aterhitta (vb)	→	återfinna	→			återfinna (vb)
d.	ivirbygning (nn)	→	överbyggnad	→			överbyggnad (nn)
e.	ighul (nn)	→	igelkott	→			igelkott (nn)
f.	sundrisker (adj)	→	söderländsk	⤳			sörländsk (adj)
g.	hvitklädder (adj)	→	hvitklädd	⤳			vitklädd (adj)
h.	tyghi (nn)	→	vitnesbörd	⤳			vittnesbörd (nn)
i.	kokiöt (nn)	→	kokött	⤳			*kokott (nn)

Figure 2: Links connecting OSwedish lemmas to Saldo

(see below). We further extend this link-up by allowing links to be created directly between OSwedish and Saldo, whenever there is no match in the Diapivot (see Figure 2c-e). Finally, the yet unlinked lemmas are matched against Saldo entries using spelling normalisation. Using techniques from Adesam et al. (2012), we extracted spelling variation rules from the Diapivot, capturing differences between Modern and Contemporary Swedish. Example of (correct and incorrect) matches found this way is found in Figure 2f-i. A fuzzy matching like this obviously introduces more errors to the link-up and methods to minimize these are part of our future work.

At this point, we only give a very limited pilot evaluation, based on 150 randomly chosen pairs of linked lemmas (see Figure 4). We mark a link as correct whenever the given Saldo entry corresponds to the definition in the OSwedish lexica. For the Diapivot-matching and the fuzzy matching we also accept broad matches, which, for the first case, means that a lemma might link to a hypernym and in the latter means that adverbs may link to adjectives, cf. *plötsligt* 'suddenly' → *plötslig* 'sudden'. The erroneous links correspond to cases when the spelling normalisation were too generous, when the extraction of a definition failed to identify the relevant parts of information, or when the linking could not identify the correct modern lemma from a set of homographs. Figure 3 shows the number of created links per word class. Black marks the links created by using the Diapivot, gray the links created directly from Old to Contemporary Swedish, and white the links created using fuzzy matching.
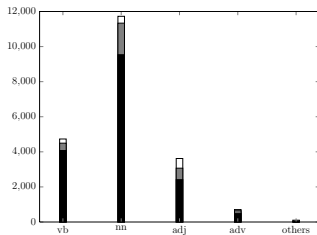

Figure 3: Number of linked lemmas per word class.

	Diapivot	Direct	Fuzzy
Correct	86%	86%	82%
Erroneous	14%	14 %	18%

Figure 4: Correctness of the different methods

3.1 Improving the link-up

An important step to increase the coverage as well as the accuracy of the link-up is to improve the quality of the digitized lexica. In their present form, inconsistent mark-up sometimes hinders us from identifying the correct definition. We are also working on refining the method for extracting the most relevant parts of a definition, to avoid mistaking lexical terms and abbreviations for being part of the real definition. A third improvement is to identify entries of the OSwedish lexica that are defined by referring to another entry within the same lexicon. This is quite common, and by interlinking these pairs and letting them share their connections to Saldo we could increase the coverage of the link-up. In Borin and Forsberg (2011), some links between Saldo and Dalin are created using a basic analysis of unknown lemmas in Dalin. For instance, unknown compounds may be linked by their head word only. This is the source of the broad matches mentioned above, and the approach would be helpful also in our case.

We further need to consider how entries with multiple matches in Saldo should be handled, and to what degree the linking should be carried out on a level of senses rather than lemmas. The semantic difference between two senses of one lemma in the OSwedish lexica may be rather big, cf. the lemma *leka* which apart from *spela* 'play' is defined as both *strida* 'fight' and *roa sig* 'have fun'. For comparison, the corresponding modern word *spela* is divided into six separate entries in Saldo. Possible reasons for this difference may be a higher number of homonyms in OSwedish, combined with a tendency of the composers of the old lexica to distinguish between many contextual nuances within one entry.

4 Tracking lexical change in large-scale material

We focus our case-study on a delimited sets of links, considering only lemmas that appear both in the Old and in the Modern or Contemporary lexica, that is links that were found without using information from the definitions. This intuitively reflects the sets of words that are etymologically related. At the current stage, we further only consider exact matches leaving out lemmas that were linked using spelling variation. While many of the lemmas meeting these criteria are assigned the same set word classes by all lexica, we chose to closer examine the cases when the information about word class differs in one of the resources. In this way we are able to track potential cases of semantic change and grammaticalization. For the rest of this paper, we will take a slightly simplified position in defining and distinguishing between content words and (grammatical) function words, seeing verbs, nouns, adjectives and adverbs (of manner, place and time) as content words, belonging to open word classes, while the other word classes, mainly prepositions, conjunctions, pronouns, sentence adverbs and particles, are considered as functions words (closed class items).

The words that the historical and modern lexica assign to different word classes are mainly interesting as illustrating cases of narrowing and broadening of meaning. The most common case is narrowing: when OSwedish verbs and nouns have several different meanings that are no longer present in the MCSwedish. Examples found in our listing are the verb *rykta* 'manage, nurture, see to, acquire' with the modern counterpart *rykta* 'see to' and the noun *gift* translating to 'marriage' and other meanings related to different gifts, *gåvor*, whose MCSwedish homonyms only means 'marriage' but also the unrelated 'poison'. Cases of semantic generalization also appear to be present in the results of our linking. For instance the noun *flykt* 'flight', only noted as 'bird travel' in OSwedish is far more polysemic in MCSwedish. Another semantic change, namely pejoration of meaning, is seen in the OSwedish noun *tukt* with definitions 'upbringing, courtesy' (positive connotation), and its MCSwedish counterpart meaning 'upbringing,

discipline' (negative connotation).

Table 1 shows the part-of-speech (mis)matches between the lexical resources. To be a candidate in this table, a part-of-speech assigned in one of the lexica have to be lacking in the other. Hence, cases where a lemma is noted as both a verb and a noun in all lexica are not included. Table 1 then shows us the number of overall matches, note however that these not necessarily show a semantic relation. Totally 109 lemma matches are shown in the table. As a first step we

Söderwall	Dalin/Saldo	counts	examples
adj/adv	noun	36	helg, kön, fat
noun	adj/adv	18	arm, hel, skum,
verb	noun	16	glosa, sena
noun	verb	10	bleka, känna,
adj/adv	prep/conj	11	vid, hos, innan, än
adj/adv	verb	7	heta, svara
prep/conj	noun	6	enka, bak, ok
noun	prep/conj	4	mot, und
prep/conj	adj/adv	1	gen

Table 1: Part of speech: Linking matches

may distinguish some patterns. Not all of the links show an (obvious) etymologically relevant connection. For instance, the apparent change from adjective/adverb to noun of some words are rather arbitrary and the high count for this reflect pairs of so called false friends. Examples are OSwedish *helg* (adverb) 'holy' and MC *helg* (noun) 'weekend'. *helg* is a spelling variant to the lemma *helagher* 'holy' in OSwedish. In Modern and Contemporary Swedish the lemma is *helig* and both part-of-speech is thus still productive; the meanings 'holy' and 'weekend' show a clear relation in that weekend correspond to 'holiday(s)'. A similar example is the OSwedish variant *kön* 'proficient' (cf. *kyn* 'sex') and MCSwedish *kön* nn 'sex'. The lemma *kyn* 'sex' is a spelling variant of *kön* in OSwedish and thus ends outside our linking experiment.

Other examples that are not likely to be considered as etymologically related lemmas are the OSwedish preposition *bi* 'by' and conjunction *ok* 'and' and their nominal homonyms in MCswedish; *bi* 'insect' and *ok* 'tool made of wood'. However a great majority of the examples in Table 1 do show a clear relation, mainly referring to contiguity in context (metonyms) or similarity (metaphors).

4.1 Semantic change inside the content word system

The most common matches we find that clearly correspond to relevant cases of semantic change are the transformation from nouns to adjectives and adverbials. Nouns for visible phenomena in OSwedish, such as *blek* 'shine' and *skum* 'twilight' and their adjective homonyms in MCswedish - *blek* 'pale' and *skum* 'dusky' - are clearly related. They correspond to well known word formations from nouns to adjectives (Ljunggren, 1939). Two similar examples that show a more unclear semantic relation are the OSwedish nouns *len* 'hill' and *ny* 'new moon'. Their adjective homonyms in MCSwedish *len* 'soft' and *ny* 'new' are not present in the OSwedish resources. That a meaning as 'new moon' had give rise to the adjective 'new', a kind of generalization, is likely but a closer look at the lemma *len* makes clear that the counterpart *lin* 'soft' is present in OSwedish. Hence the variants *len* 'incline' and MCSwedish *len* 'soft' are hardly related lemmas.

Other interesting semantic correlations are those between nouns and verbs. The Old Swedish

nouns *bleka* 'chalk' and *känna* 'knowledge' should less or more (in that order) be related to the MCSwedish meanings of 'to bleach' (i.e. use chalk, 'kalk' to bleach) and 'to feel' (i.e. some relation between mental and physical states). Interesting changes in the opposite direction are verbs that have nominal counterparts in newer forms of Swedish, such as *glosa* 'interpret' and *sena* 'delay'. The nominal counterpart in MCSwedish *glosa* 'vocable to learn' is clearly related to the action 'interpret'. In the first sight we might imagine a relation even between the verb *sena* 'delay' and the body part *sena* 'sinew' given Dalin's definition, "shiny, round or flat, rubbery strings, through which the muscles are attached". None of the lexical resources discuss this possibility, and a metaphorical relation between delay and rubbery (strings) is of course forced and not likely to be apparent. There are no other matches between verbs and other word classes. It may be due to the fact that the most common change of verbs is the rise of auxiliaries. No such relation is identified, since we do not have enough information on potential auxiliaries in more than one of the resources (Dalin).

Furthermore, we are able to identify some OSwedish adverbs and adjectives with only verbal counterparts in MCSwedish. Exampels are *heta* 'hot', and *svara* 'violent'. However the MCSwedish verbs *heta* 'to name' and *svara* 'to answer' are present in both OSwedish and MCSwedish, which implies that the adjectives in OSwedish are probably an outcome of the more general polysemic nature of OSwedish lemmas, discussed in Section 2. More interesting are the OSwedish adjective as *rena* and *ren*, 'completely', 'honest' which may have given rise to an verbal action *rena* 'to make something clean' (read 'to make something better or more complete, hence more honest'), a possible metaphorical change even though untypical in going from abstract to concrete.

4.2 Cases of grammaticalization

Table 1 also shows us potential examples of grammaticalization. The most striking example of grammaticalization found in our experiment are well known, i.e. the development of a preposition *mot* 'to(wards)' from the OSwedish noun *mot* 'meeting'. The bridging contexts are likely related to the use *om ganga/koma a mot*, 'walk/come to meet', in which the head noun come to be replaced with only the directional meaning. Other examples that might be defined as grammaticalization is the prepositional use of *hos* 'at' and *vid* 'by' in MCSwedish. In OSwedish they are also used as adverbs. Söderwall gives the adverbial meanings for *hos* 'home at' and for *vid* 'at that time, due'. As regards *hos* it is also likely that the noun *hus* 'house' forms the origin of the adverbial use (Hellquist, 1957), although this connection could not be directly identified in our experiment. Even if those words also are used as prepositions in OSwedish, the restricted use to prepositions in MCSwedish indicates a grammaticalization process. Likewise, some adverbs and adjectives have conjunctional counterparts in MCSwedish. Examples are *innan* 'inside, during', *fast* 'fixed' and their conjunctional counterparts *innan* 'until' and *fast* 'though'. They are not noted as conjunctions in Söderwall and we may see this as a grammaticalization process triggered by the conceptual metaphorical change from positions in space to a fixed position in time in the case of *innan*, and possibly from space/time to concessive meaning in the case of *fast* (c.f. Traugott and Dasher, 2002).

The reverse process, called degrammaticalization (discussed and problematized elsewhere, e.g. Andersson, 2008), might be apparent when first looking at the OSwedish preposition *gen* 'against', 'towards', that only has an adjective as counterpart in MCSwedish *gen* 'near'. However, our linking approach has once again obscured the presence of a closer OSwedish counterpart, this time *gin* 'near'. The etymological relationship between the adjective and prepositional

meanings seem to be very fuzzy, (Hellquist, 1957). Other prepositions and conjunctions in OSwedish such as *ok* 'and', *bi* 'at' are discussed above and not likely related to their homonymic nominal counterparts. Another case is the OSwedish preposition *bak* 'behind' that has a nominal homonym in MC Swedish *bak* 'back, spine'. However the latter meaning has the OSwedish counterpart *baker* which with great certainty is the origin with respect to typological and etymological data (Hellquist, 1957). Another interesting example in our linking experiment, OSwedish adjective *enka* 'single, a few' is also noted as a pronoun in Söderwall with the unclear meaning *någon* 'someone'. The MCSwedish counterpart *änka* 'widdow' is likely related to the adjective (ie. 'single') but more doubtful to the pronoun given Söderwall's unsure position of it's pronominal use. Thus, it has to be defined as a change inside the content word system.

5 Conclusion and Outlook

We have introduced a first version of a diachronic lexical resource, connecting three OSwedish lexica to the contemporary resource Saldo, and consequently to the Swedish FrameNet++. To be able to entirely connect the lexica, a considerably higher amount of work would be needed. However, we believe that the results we got by linking up the lexica with a fully automated process will already be of use in analysing historical texts and for processes of change.

Based on the links we found between the historical and modern lexica, we preformed a small study showing some interesting cases of more or less transparent lexical change. Semantic changes inside the content word system such as generalization, specialization, metonyms and metaphors have been identified and further discussed. Potential cases of grammaticalization have also been identified. The most clear cases were the few examples of nouns or adverbs (of space and time) that come to be less or more restricted to grammatical words in MCSwedish. Examples thereof are *mot*, *hos* and *vid*, all of them being used as content words in OSwedish. A few cases of potential degrammaticalization have further been discussed. However, to reach a more solid and deep analysis of the relations identified, it is necessary to complement this study with corpora research.

6 Future work

In Section 3.1 we gave an overview of the tasks at hand for improving the quality and coverage of the link-up of Old Swedish and Contemporary Swedish lemmas. In doing so we may identify several interesting cases of potential lexical change on different levels. For example, a more detailed analysis of the material studied in Section 4 might discover many semantic micro-changes, mainly inside the content word system. Attempts to carry out the linking on a level of senses, as discussed in Section 3, would be highly interesting to evaluate, with tracking conceptual-semantic domains as a possible aim. A related idea is to identify other information in the lexical entries, for example more detailed notes on the use of words, e.g. "obsolete", would be highly interesting to study. Alongside those goals more lexical resources in digitalized form is desirable in linking attempts, such as the already digitalized versions of Swedberg's (Swedberg, 2009) and F.A. Dahlgren's dictionaries (Dahlgren, 1960) on Modern Swedish.

Acknowledgments

References

Adesam, Y., Ahlberg, M., and Bouma, G. (2012). *bokstaffua, bokstaffwa, bokstafwa, bokstaua, bokstawa. . .* Towards lexical link-up for a corpus of Old Swedish. In *Proceedings of LTHist 2012*.

Andersson, P. (2007). *Modalitet och förändring. En studie av må och kunna i fornsvenska*. Göteborgsstudier i nordisk språkvetenskap 10, Institutionen för svenska språket, University of Gothenburg, Gothenburg.

Andersson, P. (2008). Swedish må and the degrammaticalization debate. In Seoane, E. and López-Couso, M. J., editors, *Theoretical and empirical issues in grammaticalization*, Typological studies in language, pages 15–32, Amsterdam/Philadelphia.

Borin, L., Dannélls, D., Forsberg, M., Toporowska Gronostaj, M., and Kokkinakis, D. (2010). The past meets the present in swedish FrameNet++. In *14th EURALEX International Congress*, pages 269–281.

Borin, L. and Forsberg, M. (2011). A diachronic computational lexical resource for 800 years of Swedish. In *Language technology for cultural heritage*, pages 41–61. Springer, Berlin.

Borin, L., Forsberg, M., and Lönngren, L. (2008). SALDO 1.0 (Svenskt associationslexikon version 2). Språkbanken, Göteborgs universitet.

Dahlgren, F. A. (1960). *Glossarium öfver föråldrade eller ovanliga ord och talesätt i svenska språket från och med 1500-talets andra årtionde*. Atelier Elektra, Köpenhamn.

Dalin, A. F. (1853/1855). *Ordbok öfver svenska språket*, volume I–II. Stockholm, Sweden.

de Melo, G. and Weikum, G. (2009). Towards a universal wordnet by learning from combined evidence. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM 2009)*, pages 513–522, New York, NY, USA. ACM.

Eriksson, M. (1992). Ett fall av grammatikalisering i modern svenska. Ba i ungdomars talspråk. *FUMS rapport 166*. Institutionen för nordiska språk, Uppsala.

Ernst-Gerlach, A. and Fuhr, N. (2007). Retrieval in text collections with historic spelling using linguistic and spelling variants. In *Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries*, pages 333–341. ACM.

Gotscharek, A., Reffle, U., Ringlstetter, C., Schulz, K. U., and Neumann, A. (2011). Towards information retrieval on historical document collections: the role of matching procedures and special lexica. *International journal on document analysis and recognition*, 14(2):159–171.

Gurevych, I., Eckle-Kohler, J., Hartmann, S., Matuschek, M., Meyer, C. M., and Wirth, C. (2012). UBY - a large-scale unified lexical-semantic resource based on LMF. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2012)*, pages 580–590.

Hellquist, E. (1957). *Svensk etymologisk ordbok 1&2*, volume 3. C.W.K Gleerups förlag, Lund.

Hopper, P. and Traugott, E. C. (2003). *Grammaticalization*. Cambridge university press, Cambridge, second edition.

Koolen, M., Adriaans, F., Kamps, J., and De Rijke, M. (2006). A cross-language approach to historic document retrieval. *Advances in Information Retrieval*, pages 407–419.

Ljunggren, K. (1939). *Adjektivering av substantiv i svenskan*. C.W.K Gleerups förlag, Lund.

Navigli, R. and Ponzetto, S. P. (2012). BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217–250.

Pumfrey, Stephen, P. R. and Mariani, J. (2012). Experiments in 17th century English: manual versus automatic conceptual history. *Literary and Linguistic Computing*. Oxford University Press.

Schlyter, C. J. (1887). *Ordbok till Samlingen af Sweriges Gamla Lagar*, volume 13 of *Saml. af Sweriges Gamla Lagar*. Lund, Sweden.

Swedberg, J. (2009). *Swensk ordabok*. Uppsala, Sweden.

Söderwall, K. F. (1884). *Ordbok Öfver svenska medeltids-språket. Supplement*. Lund, Sweden.

Söderwall, K. F. (1953). *Ordbok Öfver svenska medeltids-språket. Supplement*, volume IV–V. Lund, Sweden.

Traugott, E. C. (2001). Legitimate counterexamples to unidirectionality. `stanford.edu/~traugott/papers/Freiburg.Unidirect.pdf`. Paper presented at Freiburg university.

Traugott, E. C. and Dasher, R. B. (2002). *Regularity in semantic change*. Cambridge university press, Cambridge.