

Doing historical linguistics using contemporary data

Øivin Andersen

University of Bergen

Abstract

Retrieving linguistic data from earlier stages of languages is a notoriously difficult task. Using large electronic corpora combined with data on frequency this task can to some extent be solved. In this article I focus on the use of token frequency as described in functional Grammaticalization Theory. Deverbal nouns are non-prototypical members of the noun class. As they get older they tend to develop into more prototypical nouns. In Grammaticalization Theory this process is called lexicalization. This was tested on some zero suffix nouns in the Norwegian newspaper corpus in 2004 using modern texts only. In this article I test these findings using older texts from the same corpus.

Key words: deverbal nouns, diachrony, lexicalization, token frequency, functional linguistics.

1. Introduction

For many languages throughout the world we have no access to historical information. As Grammaticalization and Lexicalization Theories have pointed out, token frequency of lexical units in a language may carry information about their history. Using frequency data from

large electronic corpora like the Norwegian Newspaper Corpus we can catch a glimpse of past developments where historical linguistic data are almost absent.¹

In this article I will draw on data from the Norwegian Newspaper Corpus to give some support to my hypothesis that deverbal nouns (henceforth DNs) tend to develop into prototypical nouns the older they get. A DN is a noun which is derived from a corresponding verb. The DN *ødeleggelse* “destruction” is derived from the verb *ødelegge* “destroy” with the addition of the derivational suffix *-else*. Following basic principles from stage theory in grammaticalization theory I postulated 6 stages in this development in an earlier article (Andersen 2007a), using data from 2004.

2. Stage theory

The process of nominalization has been described as a process of decategorialization (Malchukov 2004:9). Several researchers have listed various operations involved in the process (Lehmann 1988, Givón 1990, Croft 1991, Dik 1997). Givón, for example, mentions that when verbs acquire nominal form, verbal agreement, tense, aspect and mode marking are either absent or severely restricted, case marking of the subject and object is modified most commonly to the genitive, and various determiners may be added modifying the deverbal noun, among others. A crucial question is whether there is any ordering of features acquired and lost in nominalization; see the typological literature for some proposals (Comrie & Thompson 1985, Noonan 1985, Lehmann 1988, Mackenzie 1987, Croft 1991, Koptjevskaja-Tamm 1993, Dik 1997).

¹ The relevance of the relationship between the frequency of lexical items and their history is also highlighted outside grammaticalization and lexicalization. Consider, for instance, Zipf's (1949) Principle of Economical Specialisation, whereby the age of a lexical item in the language correlates inversely with its size and directly with its frequency.

Some studies (like Comrie & Thompson 1985) involve frequency considerations. In particular, aspect and voice may be retained in nominalization, tense rarely so, and mood and verbal agreement virtually never. Lehmann's Desententialization Scale of decategorialization (Lehmann 1988, cited in Malchukov 2004:11) is quite elaborate (> represents a 'prior to' relation):

- (a) Constraints on/loss of illocutionary elements > constraints on/loss of mood/modal elements > constraints on/loss of tense and aspect > dispensability of complements loss of personal conjugation/conversion of subject into oblique > no polarity > conversion of verbal into nominal government > dispensability of subject/constraints on complements
- (b) Combinability with adposition/case affix

Loss of sentential properties on the part of the verb and the increasing nominality endow it with distributional properties of a noun, such as combinability with adpositions or case affixes. As Mackenzie (1987) has pointed out, nominalization involves valency reduction.

This is also an important point in the development of the Norwegian deverbal nouns.

These stages were based on purely synchronic data from the corpus. In this investigation I use supplementary data consisting of older newspaper texts collected by Uni Digital and used for comparative reference. Data support for a hypothesis like this would normally require two mutually independent empirical domains: The corpus domain where synchronic data are drawn from the corpus based on token frequency on the one hand and an authentic diachronic data domain on the other hand. The basis for my postulation of stages was, however, based on the first domain only. In this article I will report on corpus data dating back to 1900 in order to investigate whether these data can give further support to my hypothesis.

3. Deverbal nouns, prototypes and diachronic paths

Recent typological research has shown that DNs in various languages tend to develop diachronically into what Vendler (1967) called perfect nouns. Relevant research on nominalization has been carried out by several researchers like Alexiadou (2001), Anderson (1985), Cowie (2000), Comrie & Thompson (1985), Hansen & Heltoft (1994), Grimshaw (1990) and Koptjevskaja-Tamm (1993). Norwegian relevant studies are Faarlund, Lie & Vennebo (2002), Kinn (1994), Lødrup (1989), Sakshaug (1999) and Vinje (1973).²

The mechanisms involved in the historical development of grammatical and lexical phenomena have been studied both in a formal generative theoretical framework (van Kemenade & Vincent 1997) and in a cognitive functional framework (Kellerman & Morrissey 1992). The latter type of framework will be adopted here.

In the functional tradition there has been an ongoing discussion on the status and validity of Grammaticalization Theory and the hypothesis of unidirectionality. When lexical elements become grammaticalized they will tend to have a considerable rise in token frequency (Hopper & Traugott 2003:124ff). This rise seems to signal linguistic changes such as semantic fading, phonological reduction, positional fixing, increased compositionality and reanalysis of word boundaries. Lexicalization, however, tends to result in semantic enrichment and reduction/loss of compositionality (Brinton & Traugott 2005:32ff).

Prototype theory is one of the most central theories in cognitive functional linguistics. Working with mental lexical structures in cognitive psychology, Eleanor Rosch found several

² For more details see Andersen 2007a.

lexical features exhibiting prototype structures (Rosch & Lloyd 1978). For the category of birds a prototype feature is the ability to fly. An exemplar like a robin will have this feature, being a prototypical member of the category of birds, whereas an ostrich, sharing important properties with the robin, like laying eggs, will lack the property of being able to fly. Hence a robin is a more prototypical representative member of the bird category than an ostrich.

Prototype theory was introduced into linguistics by amongst others Taylor (1989).³ A central point in prototype theory is the notion that some members of a linguistic category are better and more prototypical representatives (members) of the category than others. Applied to word classes this means that some verbs are more prototypical (i.e. more central members of the class) than others. Likewise, some nouns are more prototypical, i.e. more nounlike than others (Hopper & Thompson 1985). This means that the categories of nouns and verbs can be defined as having characteristic features where some features are absent in the non-prototypical members (lexemes). In Vendler's terms a prototypical noun would be a perfect noun and a non-prototypical noun would be an imperfect noun. An imperfect noun is rather colourfully defined as "a noun with a live and kicking verb inside itself" (Vendler 1967:131).

Hopper & Thompson (1985) and Taylor (2004) see categoriality as a gradient phenomenon. Here it is important to add that categorical gradience does not imply fuzzy concept borders. As I have demonstrated (Andersen 2007b) the border between nouns and verbs is discrete even though both word classes do have distinct denotational gradience.

My central hypothesis is that DNs follow a diachronic path leading from imperfect noun to perfect noun, i.e. from a non-prototypical member of the noun class in the direction of a more

³ Geeraerts (1992) gives a nice survey of diachronic semantics within this framework.

prototypical member of the class. Prototype properties both at the morphosyntactic level and on the semantic level may be cancelled in different contexts of use, as Taylor (2004), Givón (1979) and Ross (2004) have pointed out. In this article I will focus on some basic properties that will be applied as the basis for my postulation of diachronic stages.

When a new DN is coined it inherits event process meaning from the corresponding verb.

Coining of new DNs typically occurs in technical and scientific writing.⁴ Halliday & Martin (1993) and Banks (2004) have investigated the historical development of DNs in Newton's scientific writings. Events and processes are most typically referred to by DNs. These DNs tend to become scientific terms with a very specific event process meaning. In other words, process event meaning is a prototypical property of the verb and is carried over in deverbal nominalization. This property is non-prototypical for nouns.

Whereas process meaning is typical for verbs, result meaning is typical for nouns. The result of a process is a static state associated with a prototypical noun property like time-stability (Givón 1979, 1984). Givón's Time Stability Hypothesis maintains that different parts of speech show gradience as to the ability to denote permanence or stability over time. Verbs like *å skyte* "to shoot" tend prototypically to be less time stable than nouns like *bygning* "building". So verbs tend to be less time stable than nouns, and adjectives tend to take a mid position between verbs and nouns (Givón 1984:55). Result meaning very often gives rise to time stable entity meaning, which is typical of nouns. As Grimshaw (1990) has pointed out, many DNs have a polysemy which is characteristic for many DNs in several languages containing process and result meaning. A DN like *tegning* 'drawing' may refer to the process of drawing something, or to the result of that process, i.e. a drawing hanging on the wall.

⁴ For discussion and examples see Andersen 2007c

Pustejovsky (1998) calls this logical polysemy. Pustejovsky makes a distinction between contrastive ambiguity (usually referred to as homonymy) and complementary polysemy. The latter type is category-preserving. Logical polysemy is defined as a systematic type of complementary polysemy where there is no change in lexical category, and the multiple senses of the word have overlapping, dependent and shared meanings (Pustejovsky 1998:28). These systematically related meanings tend to include large sets of nouns. A typical example is the figure/ground reversals discussed in Pustejovsky (1998:31f.).

As Bybee (1985:33ff) has demonstrated, some prototypical grammatical verbal properties are more easily lost in nominalization than others. Her basic principle is that the more relevant the category is to the root meaning of the verb, the more is it likely to be retained, i.e. carried over in nominalization. Aspect is more relevant to the root than tense, so aspect distinctions are retained to a greater extent than tense. In Norwegian there is a closed set of zero suffix DNs referring to sound and movement which have an aspectual distinction of imperfect meaning (for the high frequent DNs ending in *-ing*) and perfect or semelfactive meaning for the corresponding zero suffix DNs. Semelfactive verbs express an instantaneous event. Examples are *sparkning* (“the phenomenon of kicking”) vs *spark*, (an instance of *kicking*) (Faarlund, Lie & Vennebo 2002: 124). As Taylor (2004: 306) has pointed out, instantiations are typical of nouns, using for instance determiners like *this* or numerals like *one*. Determiners and numerals are seldom used with *-ing* DNs with imperfect meaning, but occur frequently with zero suffix nouns with perfective meaning.

Another important difference between verbs and nouns is the fact that verbs typically have syntactic argument structure and semantic participant structure associated with it, whereas

nouns typically lack this property (Chomsky 1970). Chomsky's own example is the verb *destroy*:

- (1) The enemy destroyed the city.

In nominalization the direct object is carried over and the subject may occur in the possessive:

- (2) the enemy's destruction of the city.

Following Grimshaw (1990) the preposition *of* is the marker of the fact that the direct object has been carried over. She calls it a theta transmitter (Grimshaw 1990:73). In Norwegian the corresponding preposition is *av*. Lehmann (1995 [1982]:2f.) calls this a grammatical use of the preposition. Normally prepositions have lexical (often local and temporal) functions.

Examples are prepositions like *on*, *at*, *behind* and *in*. The lexical prepositional phrases function typically as adjuncts, i.e. constituents outside the role frame of the corresponding verb. In DNs with transferred argument structure adjuncts are unchanged in nominalization, as in *servere i restauranten* 'serve in the restaurant', *servering i restauranten* 'serving in the restaurant'.

Prepositions occurring with DNs in postnominal position inside the role frame of the corresponding verb are realized as two different types of prepositional phrase dependents. On the semantic level, these are known as participants and non-participants. Participants are selected by the meaning of the head DN or the main verb, whereas non-participants are outside the selectional frame of the head DN or the main verb. On the syntactic level these correspond to the distinction between complements and non-complements (Saeed

2003:148ff). Complements are subject to the subcategorization restrictions of the head DN or the main verb, whereas non-complements are not subcategorized by the head noun or the main verb. Some examples from the newspaper corpus may serve to illustrate this:

- (3) *Liv Grete presterte å treffe med tre skudd-Ø på første skyting, men ble stoppet...*

Liv Grete managed to hit with three shot-NMLZ on first shooting, but was stopped

"Liv Grete managed to hit with three shots on the first shooting, but was stopped."

(AP0315)⁵

- (4) *...Middlesbrough ikke maktet et eneste skudd-Ø på mål i løpet av kampens 90 minutter.*
...Middlesbrough not power one single shot-NMLZ on goal during match's 90 minutes

"Middlesbrough did not have a single shot on the goal during the 90 minutes of the match".

DB40925

In (3) the prepositional phrase *på første skyting* is an adverbial referring to place. It is not related to the lexical meaning of the DN *skudd*. Hence it is a non-participant. In (4) the prepositional phrase *på mål* is related to the lexical meaning of *skudd* and hence a participant in the event denoted by *skudd*.

I also distinguish between two types of complements: arguments and non-arguments.

Arguments are those complements which grammatically code (or license) functional elements as direct objects both on sentence level in the case of main verbs, e.g. *Læreren løste problemet i klasserommet* 'The teacher solved the problem in the classroom') and on pharse level in the case of DNs, using the grammatical preposition *av* "of", e.g. *Lærerens løsning av problemet tok lang tid* 'The teacher's solving of the problem took a long time'".

⁵ The codes after the examples refer to name of newspaper and date of issue, cf. <http://avis.uib.no/om-aviskorpuset/innhold>. Thus, AP refers to the newspaper Aftenposten, and the date is 15th of March. The examples are taken from the tagged 2004 part of the corpus.

Within the DNs, there is a further distinction between those which code participants as arguments (i.e. *Lærerens løsning av problemet tok lang tid*), using grammatical coding (the grammatical preposition *av* “of”) and those which code participants as non-arguments (using a lexical preposition), e.g. *Lærerens løsning på problemet lå på bordet* “The teacher’s solution to the problem was on the table”). So the DN *løsning* has both process and result meaning. The DN *løsing* has an unambiguous process meaning. In some restricted pairs of words we have an overt distinction between DNs ending in *-ing* (having process meaning only (as in *bygging* “the process of building something”) and DNs ending in *-ning*, as in *bygning* (the result of the process of building something).

4. Methods and data

4.1. Using a synchronic corpus only

Based on relevant elements of Prototype theory and Grammaticalization Theory (as outlined in the introduction) I postulated 6 stages for the historical development of DNs in Norwegian. I call these the prototype cline of zero suffix DNs:⁶

1. Establishment of event process meaning and reference
2. Development of result meaning, specific reference, development of logical polysemy by reanalysis and analogical spreading
3. Development of idiosyncratic polysemy and increased semantic distance from the verb
4. Development of idiosyncratic morphophonological alternation and increased form distance from the verb

⁶ For a more detailed description of the stages I refer to my article (Andersen 2007a:66ff).

5. Loss of aspect distinctions by analogical spreading
6. Loss of prepositional dependent and lexical differentiation

Malchukov (2004:9) describes the historical process of nominalization as a process of decategorizaton, i.e. the gradual loss of verbal properties. This process can also be viewed as a movement from the periphery members of the noun class in the direction of the nucleus members of the same class.

But as Grammaticalization Theory points out, these stages are not to be seen as consecutive and well defined thresholds of development, i.e. as mutually exclusive time periods. In a comprehensive process like deverbal nominalization involving several different morphological types (for a survey, see Andersen 2007c), different types may not only represent different stages, but different members of the same morphological type may be on different stages of development, as we shall see. Old forms of lexemes may persist for an unpredictable period of time. These forms may live side by side along with newer forms for a long time. This may lead to what Hopper & Traugott (2003:125ff) call layering, i.e. variability of analogical old and new forms.

In most cases productivity, frequency and compositionality are closely correlated phenomena. In Grammaticalization Theory it has long been recognized that frequency data are important in showing the unidirectionality of how lexical forms move into grammatical roles. Two basic types of frequency are distinguished: type frequency and token frequency. Type frequency refers to the number of items that are available to a particular class of forms (Hopper & Traugott 2003:124ff.). For instance, the number of lexical items taking the deverbal nominal suffix *-ing* is very high (comprising most

DNs in Norwegian), whereas the number of lexical items taking the deverbal nominal suffix *-t* (as in *drive* “to manage” vs. *drift* “management”) is much lower.

But, as Hopper & Traugott (2003:125) have pointed out, most attention has been paid to token frequency, i.e. the number of times a particular form occurs in texts or the changes in frequency of forms or constructions over time. These frequency data are very characteristic of the grammaticalization of grammatical formsToken frequency means the number of times a particular form exemplar occurs in texts, or, the changes of frequency of form exemplars or constructions exemplars over time. A marked increase of token frequency may indicate grammaticalization. Conversely a marked decrease may indicate lexicalization.

Changes such as semantic fading (called bleaching in Grammaticalization Theory), phonological reduction, positional fixing, increasing compositionality and erasure of word boundaries, are inseparable from the absolute token frequency of the forms and the frequency with which they occur with other forms. In the opposite process, lexicalization, factors such as semantic enrichment and reduction of compositionality will be assumed to have a close connection to token frequency.

But the type frequency of the different morphological types is also an important indication of the status of the DN. Nouns with high type frequency, such as the *-ing* nominals in Norwegian, tend to retain many of their verbal characteristics (such as process meaning and grammatical argument structure), whereas nouns with low type frequency tend to take on noun characteristics (such as result meaning, loss of grammatically coded argument structure

and development of specific reference and use with determiners). Thus, the development of DNs can be regarded as a process of lexicalization (Brinton & Traugott 2005).

Data on type frequency are very difficult to extract automatically from an electronic corpus. This is especially the case for zero suffix nouns and some low frequency types such as *-t* (as in *drift* “continuous production”) in Norwegian. An investigation on type frequency in technical texts revealed that zero suffix DNs had a much higher type frequency than expected, and seem to enter into alternations with the corresponding high productive *-ing* DNs (Andersen 2007c: 159). I investigated the type frequency of the different DN suffixes in a technical manual for the Gullfaks A platform describing the steam generation system. The manual contained 150 pages. I found 154 DNs in the manual. 57,14% had the *-(n)ing* suffix, 20,13% had the zero suffix, 10,39% had the *-sjon* ending. The other suffixes ranging from 3,25% to 0,65%. The high percentage of the zero suffix nouns was surprising.

Examples from this manual showed examples like *start av maskiner* (“start up of engines”) alternated with *starting av maskiner* (with the same meaning) (Andersen 2007c:159).

The *-ing* vs. zero suffix alternations did not exhibit the perfect vs. imperfect meaning alternations mentioned above (cf. the example *sparking* “the phenomenon of kicking” which has imperfective aspect, versus *spark* “a kick” which has perfective or semelfactive meaning), but both members of the alternations had imperfect process event meaning. In these cases the zero suffix DNs occurred with grammatically coded arguments, using the theta transmitter *av*:

(5) standardprosedyre for **isoler-ing av** reguleringsventiler

Standard procedure for isolate-NMLZ of control valves

"standard procedure for insulation of control valves"

(6) **Sjekk-Ø av** manuelle ventiler må foregå etter kontortid

check-NMLZ of manual valves must take place after office hours.

"The testing of manual valves must take place after office hours. "

As for my data on token frequency, I have used the comprehensive Norwegian Newspaper Corpus (Andersen and Hofland, this volume). As of date the size of this corpus is about 850 million token words. At the time of my investigation the number of words were 430 million. The corpus allows automatic extraction of absolute token frequency and relative frequency per 100 million words.

Based solely on these synchronic data I investigated the six zero suffix DNs *skudd* "shot", *slag* "beat", *brudd* "breach, violation", *drap* "murder", *salg* "sale" and *kjøp* "purchase" and concluded that they occupy different positions on the prototype cline. The selection of these six DNs was based on the following assumptions:

1. The zero suffix nouns with perfect aspect meaning, like *skudd* "shot" and *slag* "beat" would be expected to be closer to the perfect noun (in terms of Vendler 1967) pole of the cline.
2. DNs closer to the perfect noun pole of the cline would be expected to have elements of idiosyncratic polysemy and increased semantic distance from the verb.
3. DNs closer to the perfect noun pole would be expected to have idiosyncratic morphophonological alternations and thus increased distance from the corresponding verb on

the morphosyntactic level. Based on these assumptions the expected distribution of these DNs would be as follows:

perfect noun	imperfect noun
<i>skudd</i> <i>slag</i> <i>brudd</i> <i>drap</i>	<i>salg</i> <i>kjøp</i>
idiosyncratic	predictable
no argument structure	argument structure
no participant structure	participant structure
low token frequency	high token frequency

Figure 1. Different degrees of nounhood of zero suffix nouns.

Token frequency showed that the DNs *kjøp* ‘buying’ and *salg* ‘selling, sale’ had a much higher frequency than the others, occurring with process event meaning. The difference between these two DNs and the others was especially salient when occurring with the preposition *av* ‘of’ in the function of a theta transmitter. *Skudd* ‘shot’ and *slag* ‘beat’ did not occur with this preposition at all. The DN *brudd* ‘breach, violation, fracture’ had a much lower frequency but occurred with the theta transmitter in about 50% of the cases. The DN *drap* “murder” had a higher frequency than *brudd* and about 75% of its occurrences with the theta transmitter *av* (cf. Andersen 2007a: 72).⁷ Occurrence with the preposition *på* functioning as a participant, as in *skudd på soldater* ‘shots at soldiers’ as opposed to *skudd på slagmarken* ‘shot in the battle field’, data showed that *brudd* occurred frequently with the participant preposition *på*. The DN *drap* had a lower frequency, but occurred mostly with the same preposition in the same function. The DNs *slag*, *skudd* and *salg* had only few occurrences with this preposition. However, the 6 occurrences with *salg* and the preposition *på* all showed

⁷ In the 2004 investigation the Dns *kjøp* with the preposition *av* had a relative frequency of 1817 and *salg + av* had a relative frequency of 3828 per 100 million words. In contrast the figures for the others were: *brudd+av* 30, *slag + av* 30, *skudd + av* 153 and *drap + av* 186. An investigation of the first 200 occurrences revealed that with *kjøp* and *salg* all the occurrences of the preposition *av* was that of a theta transmitter. The figures for the others were: *brudd* 16, *skudd* 0 and *slag* 0.

a different meaning of the DN compared to the corresponding verb. The collocation *salg på* (as opposed to *salg av*) means that the selling takes place at a lower price than the standard price. This added meaning component must have occurred after the coining of the DN. This seems to be a case of semantic enrichment, i.e. a characteristic of lexicalization. Enrichment relates to stage 3 of the cline where the development of idiosyncratic polysemy and semantic distance is focussed on.

infinitive	deverbal noun	predictability	frequency	argument structure	participant structure
skyte	skudd	idiosyncratic	low	no	no
slå	slag	idiosyncratic	low	no	little
bryte	brudd	idiosyncratic	low	some	full
drepe	drap	idiosyncratic	low	some	almost full
selge	salg	idiosyncratic	high	full	little
kjøpe	kjøp	predictable	high	full	little

← **Formatert:** Linjeavstand:
Enkel

Table 1. Argument and participant structure with zero suffix DNs

As Grammaticalization theory has pointed out (Hopper & Traugott 2004) grammaticalization paths seem to be organized onomasiologically, i.e. semantic bleaching must occur before phonological reduction takes place. This is iconically the most plausible sequence. According to this iconic principle an increased distance in meaning between the verb and its corresponding DN will give rise to an increase of idiosyncrasy of the relationship between the verb and the DN. This was focussed on in stage 4. As table 1 shows, the only fully predictable morphophonological alternation between the infinitive form of the verb and the DN is in *kjøpe* vs. *kjøp*. This DN has the highest token frequency and it occurs most regularly with argument structure. The alternations *bryte – brudd*, *skyte – skudd* and *slå – slag* have both

vowel alternation and an unpredictable final consonant in the DN. In *drepe-drap* and *selge – salg* there is vowel alternation only.

As regards stage 6 there seems to be a nice correlation between the status of being an imperfect noun on the one hand, and the presence of argument structure on the other hand.

To sum up, these data show that the 6 investigated DNs occupy different positions along the perfect – imperfect noun continuum.

5. Using older texts from the corpus

The next question is then: If we have access to historical data would these tendencies be further supported? As is well known by Nordic linguists there is very scant historical evidence from written records at least from the Middle Ages. Systematic relevant data in sufficient quantities are at present unavailable.

But the Norwegian Newspaper corpus contains some older texts dating back to 1900⁸. The older texts contains the following:

1. Texts from around 1900: 217.000 word tokens
2. Texts from around 1925: 355.000 word tokens
3. Texts from around 1950: 355.000 word tokens
4. Texts from 1971: 940.000 word tokens
5. Texts from three newspapers 1981: 810.000 word tokens
6. The newspaper Bergens Tidende 1995-1998: 8.621.000 word tokens

⁸ These texts are not a part of what we usually refer to as the Norwegian Newspaper Corpus, but a different set of older newspaper material collected by Uni Digital and used for comparative reference.

Item 6 contains 4 papers a month.

I studied the 6 DNs in figure 1, using KWIC-concordances to see to which extent they occurred with a prepositional dependent in the form of an argument coded by the preposition *av* or in the form of a participant coded by the preposition *på*. Occurrence without the prepositional dependent is taken as an indication of the status of the DN as a perfect noun.

In the process I had to resolve the usual disambiguation of polysemy/homonymy. Thus, the DN *slag* has meanings which are not (at least not obviously) derived from a corresponding verb, like “sort, type”, e.g. *bøker av ulike slag* “books of different types”. A different meaning appeared in the collocation *å være i godt slag* “to be in a good physical condition”. These were all excluded. Likewise with the DN *kjøp*, where the imperative form “buy” occurred in many advertisements, had to be excluded.

skudd

	occurrences	percentages
1900	0	0,000
1925	9	0,007
1950	27	0,002
1971	42	0,004
1981	12	0,001
1994-1998	298	0,003

Table 2. Older text occurrences of *skudd*

As table 2. shows the DN *skudd* has a very low overall frequency. Data revealed that it consistently occurs without a prepositional determiner. Moreover, it has a perfective, semelfactive aspectual meaning as opposed to the corresponding process denoting imperfective *-ing* DN *skyting* “shooting”, which freely occurs with argument structure. This points in the direction of *skudd* being a perfect noun.

slag

	occurrences	percentages
1900	17	0,007
1925	29	0,008
1950	31	0,008
1971	53	0,005
1981	38	0,004
1994-1998	335	0,003

Table 3. Older text occurrences of *slag*

The DN *slag* has a slightly higher overall frequency than *skudd*, but occurs consistently without propositional dependent. This also points in the direction of *slag* being a perfect noun. Although percentages are consistently low, they seem to have a falling tendency.

brudd

	occurrences	participant <i>på</i>	percentages
1900	0	0	0,00
1925	7	1	0,01
1950	14	7	0,03
1971	70	34	0,07
1981	27	12	0,03
1994-1998	452	284	0,05
2. half 94	8	6	
1. half 95	56	31	
2. half 95	55	29	
1. half 96	45	29	
2. half 96	39	25	
1. half 97	54	42	
2. half 97	52	33	
1. half 98	64	36	
2. half 98	79	53	

Table 4. Older text occurrences of *brudd*

The concordances revealed most occurrences of result meaning of *brudd*. But here there is distinct presence of prepositional dependent. The percentages do not reveal any tendency. However no arguments with *av* was detected in these texts. What complicates matters here, however, is the fact that dependents may be suppressed in contexts where they are pragmatically considered as given information or considered as our common knowledge of the world.

Example (7) and (8) may serve to illustrate this:

- (7) ...til legeundersøkelse, men slapp med mindre brudd [på x] og slagskader.
...to medical examination, but escaped with minor fractures [on x] and blow injuries.

950614

- (8) ...er omfattende. Han pådro seg alvorlige brudd på kraniet...
...are considerable. He received serious fractures on the scull...

950614

The participant in square brackets i (7) is not present in the text. The overall frequency is higher than for *skudd* and *slag*, but the frequency data from the oldest texts are uncertain. The percentages are again very low, and show no tendency. The presence of participant structure coded with *på* is conspicuously present, but again the suppression of this dependent may be due to information structure distribution in different contexts. The consistent absence of argument structure points in the direction of perfect noun status, as expected.

drap

	occurrences	argument <i>av</i>	participant <i>på</i>	percentages
1900	1	0	0	0,001
1925	3	0	1	0,001
1950	6	0	6	0,001
1971	66	2	33	0,007
1981	26	0	5	0,003
1994-1998	350	2	57	0,004
2. half 94	3	0	1	
1. half 95	60	0	8	
2. half 95	34	0	9	
1. half 96	38	1	9	
2. half 96	32	0	3	
1. half 97	43	1	5	
2. half 97	54	0	11	
1. half 98	44	0	7	
2. half 98	42	0	4	

Table 5. Older text occurrences of *drap*

The overall frequency for *drap* is in fact lower than for *brudd*. The percentages do not indicate any tendency. On the other hand some occurrences of argument structure may indicate a stronger imperfect noun status for *drap* than for *brudd*.

salg

	occurrences	argument <i>av</i>	participant <i>på</i>	percentages
1900	7	2	0	0,003
1925	23	14	0	0,006
1950	16	10	0	0,004
1971	57	40	0	0,006
1981	42	23	1	0,005
1994-1998	724	370	5	0,008
2. half 94	13	7	0	
1. half 95	85	32	1	
2. half 95	86	44	0	
1. half 96	66	38	0	
2. half 96	83	44	0	
1. half 97	102	53	3	
2. half 97	113	59	0	
1. half 98	86	41	0	
2. half 98	90	52	1	

Table 6. Older text occurrences of *salg*

As expected the overall token frequency for *salg* is much higher than for *skudd*, *slag*, *brudd* and *drap*. Again the percentages show no tendency. But the presence of argument structure is frequent. However, the special meaning of *salg* in the sense “sale at a lower price than normal price” can be detected from 1981 and onwards. It has increased its frequency since 2000 and onwards. It is an interesting fact that this meaning of *salg* never occurs with the theta transmitter *av*, but only with the participant coding preposition *på*.

kjøp

	occurrences	argument <i>av</i>	participant <i>på</i>	percentages
1900	0	0	0	0,000
1925	25	16	0	0,011
1950	11	7	0	0,003
1971	27	25	0	0,003
1981	30	15	0	0,003
1994-1998	320	216	3	0,003
2. half 94	5	3	0	
1. half 95	40	22	0	
2. half 95	46	33	0	
1. half 96	43	27	0	
2. half 96	36	24	0	
1. half 97	44	30	3	
2. half 97	32	26	0	
1. half 98	45	28	0	
2. half 98	29	23	0	

← ----- **Formatert:** Tabulatorer: 14
mm, Midstilt

Table 7. Older text occurrences of *kjøp*

The DN *kjøp* is the clearest case of an imperfect noun. It has a high overall frequency, and occurs regularly with argument structure.

6. Discussion and conclusions

The collection of older texts in the Newspaper corpus is a valuable addition to the corpus and can certainly shed new light on the development of the Norwegian language. But these corpora are too small to give further support to my hypothesis. Tables 2 to 7 indicate that these nouns are at the same stage of development as my original data from 2004 show.

This means that my hypothesis has not altered status: It has not been strengthened and it has not been weakened. As regards token frequency the oldest data are so small in quantity that very little can be concluded from them.

So, is my hypothesis really an empirical one? Is it possible in Popper's terms (Popper 1980: 32f) to conceive a method which will allow me to test this hypothesis?

Considering the fact that older diachronic data are at present unavailable the only possible solution is to make further use of the Newspaper corpus to monitor this development more closely over time. The corpus is continuously being monitored and refined, and it is my hope that this study can be pursued further in the future.

References

- Andersen, Øivin. 2007a. Deverbal nouns, lexicalization and syntactic change. *Nordic Journal of Linguistics*. 30 (1): 55-86.
- Andersen, Øivin. 2007b. Indeterminacy, context, economy and well-formedness in specialist communication. In *Indeterminacy in Terminology and LSP*. Bassey Edem Antia (ed), 3-15. Amsterdam: John Benjamins.
- Andersen, Øivin 2007c. Deverbal nouns as hybrids: Some textual aspects. In *Evidence-based LSP. Translation, Text and Terminology* [Linguistic Insights Series 47], Khurshid Ahmad Margaret Rogers (eds), 155-171. Bern:Peter Lang
- Anderson, Steven R. 1985. Typological distinctions in the word formation.
In *Language Typology and Syntactic Description*. vol. 3, 3–56. Timothy Shopen (ed). Cambridge: Cambridge University Press.
- Alexiadou, Artemis. 2001. *Functional Structure in Nominals: Nominalization and Ergativity*. Amsterdam: John Benjamins.
- Banks, David. 2004. On the historical origins of nominalized process in scientific text.
English for Specific Purposes 24.3, 347–357.
- Brinton, Laurel J. & Traugott, Elizabeth Closs. 2005. *Lexicalization and Language Change*. Cambridge: Cambridge University Press.
- Bybee, Joan. 1985. *Morphology: A Study of the Relation Between Meaning and Form*. Amsterdam: John Benjamins.
- Chomsky, Noam. 1970. Remarks on Nominalization. In *Readings in English Transformational Grammar*. A. Jacobs & P.S. Rosenbaum (eds), Massachusetts: Ginn and co., Waltham.

- Comrie, Bernard & Thompson, Sandra A. 1985. Lexical nominalization. In *Language Typology and Syntactic Description*, Timothy Shopen (ed), vol. 3, 349–398. Cambridge: Cambridge University Press.
- Cowie, Claire. 2000. The discourse motivations for neologizing: action nominalization in the history of English. In *Lexicology, Semantics and Lexicography*. Julie Coleman (ed), Philadelphia: John Benjamins, 179–207.
- Croft, William. 1991. *Syntactic Categories and Grammatical Relations: The Cognitive Organization of Information*. Chicago: University of Chicago Press.
- Dik, Simon C. 1997. *The Theory of Functional Grammar: Complex and Derived Constructions*. Berlin: Mouton de Gruyter.
- Faarlund, Jan Terje, Lie, Svein & Vannebo, Kjell Ivar. 2002 3rd edition. *Norsk referansegrammatikk* [Norwegian reference grammar]. Oslo: Universitetsforlaget.
- Geeraerts, Dirk. 1992. Prototypicallity effect in diachronic semantics: A round-up. In *Diachrony within Synchrony: Language History and Cognition*. Günther Kellermann & Michael D. Morrissey (eds), 183-205. Frankfurt am Main: Peter Lang.
- Givón, Talmy. 1979. *On Understanding Grammar*. New York: Academic Press.
- Givón, Talmy. 1984. *Syntax: A Functional-Typological Introduction*, vol. 1. Amsterdam: John Benjamins.
- Givón, Talmy. 1990. *Syntax: A Functional-Typological Introduction*, vol. 2. Amsterdam: John Benjamins.
- Grimshaw, Jane. 1990. *Argument Structure* [Linguistic Inquiry Monograph 18]. Cambridge, MA: MIT Press.
- Halliday, M. A. K. & Martin, John R. 1993. *Writing Science: Literacy and Discursive Power*. London: The Falmer Press.

- Hansen, Erik & Heltoft, Lars. 1994. Kentaurnominaler i dansk [Centaur Nominalizations in Danish]. In *NORDLEX-Prosjektet. Sammensatte substantiver i dansk*. 57–67. Irene Baron (ed.), Copenhagen: Institut for Datalingvistikk, Handelshøjskolen i København.
- Hopper, Paul J. & Thompson, Sandra A. 1985. The iconicity of the universal categories ‘noun’ and ‘verb’. In *Natural Syntax: Iconicity and Erosion*. John Haiman (ed), 151–183.
- Hopper, Paul J. & Traugott, Elizabeth Closs. 2003. *Grammaticalization*, Cambridge: Cambridge University Press.
- Kellermann, Günter & Morrissey, Michael (eds.). 1992. *Diachrony within Synchrony: Language History and Cognition*. Frankfurt am Main: Peter Lang.
- Kemenade, Ans, van & Nigel Vincent. 1997. *Parameters of Morphosyntactic Change*. Cambridge: Cambridge University Press.
- Kinn, Torodd. 1994. Prepositional phrases in Norwegian: Internal Structure and External Connections – an LFG analysis. Master thesis, University of Bergen. [Department of Linguistics and Phonetics. Skriftserie nr. 45, Serie B.]
- Koptjevska-Tamm, Maria. 1993. *Nominalizations*. London: Routledge.
- Lehmann, Christian. 1988. Towards a typology of clause linkage. In *Clause Combining in Grammar and Discourse*, John Haiman & Sandra A. Thompson, (eds.), 181–225. Amsterdam: John Benjamins.
- Koptjevska-Tamm, Maria. 1993. *Nominalizations*. London: Routledge.
- Lehmann, Christian. 1988. Towards a typology of clause linkage. In *Clause Combining in Grammar and Discourse*, John Haiman & Sandra A. Thompson (eds.), Amsterdam: John Benjamins, 181–225.
- Lehmann, Christian. 1995 [1982]. *Thoughts on Grammaticalization* [LINCOM Studies in Theoretical Linguistics 1]. Munich: LINCOM EUROPA.

- Lødrup, Helge. 1989. *Norske Hypotagmer. En LFG-beskrivelse av ikke-verbale hypotagmer* [Norwegian hypotactic constructions: a description of non-verbal hypotactic constructions]. Oslo: Novus Forlag.
- Mackenzie, Lachland. 1987. Nominalization and basic constituent ordering. In *Ins and Outs of Predication* Johan van der Auwera & Louis Goosens (eds.), 94–105, Dordrecht: Foris.
- Malchukov, Andrej L. 2004. *Nominalization/verbalization: Constraining a Typology of Transcategorial Operations*. Munich: LINCOM EUROPA.
- Noonan, Michael. 1985. Complementation. In *Language Typology and Syntactic Description, Volume II: Complex Constructions*, Timothy Shopen (ed.), 42–141. Cambridge: Cambridge University Press.
- Popper, Karl Raymond. 1980. *The Logic of Scientific Discovery*. London: Hutchinson.
- Pustejovsky, James. 1998. The *Generative Lexicon*. Cambridge, MA: MIT Press.
- Rosch, Eleanor. 1978. Principles of Categorization. In *Cognition and Categorization*, Eleanor Rosch & Barbara B. Lloyd (eds.). Reprinted in *Fuzzy Grammar. A Reader*, Baas Aarts, David Denison, Evelien Keizer & Gergana Popova (eds) 2004, 91-109. Oxford: Oxford University Press.
- Ross, John Robert. 2004. Nouniness. In *Fuzzy Grammar. A Reader*. Baas Aarts, David Denison, Evelien Keizer & Gergana Popova (eds) 351-423. Oxford: Oxford University Press.
- Saeed, John I. 2003. Semantics. 2nd edition. London: Blackwell Publishing.
- Sakshaug, Laila. 1999. Norwegian Compound Deverbal Nouns: An Autolexical Analysis in Morphology, Syntax and Semantics. Dr. artium thesis, Faculty of Arts, Department of Scandinavian Studies and Comparative Literature, University of Trondheim. Norway.

Taylor, John R. 1989. *Linguistic Categorization: Prototypes in Linguistic Theory*. Oxford: Clarendon Press.

Taylor, John R. 2004. Grammatical Categories. In *Fuzzy Grammar. A Reader*. Baas Aarts, David Denison, Evelien Keizer & Gergana Popova (eds) 293-309. Oxford: Oxford University Press.

Vendler, Zeno. 1967. *Linguistics in Philosophy*. Itacha, NY: Cornell University Press.

Vinje, Finn-Erik. 1973. Verbalsubstantiv avleddet med -else, -ing og nullsuffiks i moderne norsk, især bokmål [Deverbal nouns derived by -else, -ing and zero suffix in Modern Norwegian]. Unpublished manuscript, University of Trondheim. Norway.

Zipf, George Kingsley (1949). *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Cambridge, MA: Addison-Wesley Press.