Compositionality and deverbal nouns. Testing Chomsky's lexicalist hypothesis

Øivin Andersen

Department of Linguistic, Literary and Aesthetic Studies

University of Bergen

Summary

Scholars have different attitudes to the relationship between general linguistics and LSP. In this article I will discuss my own view on this relationship and its theoretical and methodological consequences. In the wake of this I will test Chomsky's lexicalist hypothesis on the compound deverbal noun (DN) constructions in Norwegian. More specifically, I will discuss what happens when DNs with carried over argument structure from the corresponding verb in unpacked phrases like *bygging av hus* ("the building of houses") are packed down into compounds like *husbygging* ("house building"). The two basic argument types subject and direct object will be discussed. Finally I will briefly discuss how petrification and fossilization may be studied in an LSP context.

1. Linguistics and LSP

As I have pointed out in my doctoral dissertation all domain specific LSP variants must be sublanguages which are parasitic on the general language (Andersen1998:35f). The only specific coding property distinguishing variants of LSP from LGP is the presence of specific terminology. These terms must in some sense be formed on the basis of some general mechanisms in the LGP. In other words: There are no LSPs without an encompassing, comprehensive LGP. I call this **the parasite hypothesis** of LSP, i.e. any LSP variant is parasitic on a given LGP in the sense that it presupposes the existence of it. As a consequence of this, I start an investigation on a specific linguistic construction in the general language and investigate how this construction behaves in different types of LSP afterwards. By using this procedure it will be possible to see which aspects of the construction are common to both LGP and its domain specific parasitic variants and which aspects are specific to the domains in question. In the initial general investigation various theories and models from general linguistics can be applied and tested. If they prove fruitful they can be tested again in the LSP part of the investigation.

Classical works within generative grammar have often been criticized for using constructed examples instead of genuine corpus data. This criticism is in many cases well funded and sound. However, the tendency to rely on corpus data only is problematic. It is almost impossible to find contrastive data where negative evidence is crucial. In these cases the constructed data must be used. Constructed data can in its turn give rise to revealing hypotheses which can be tested on corpus data.

Consequently, we need both constructed data and corpus data in our investigation. An initial investigation into a construction is best carried out with constructed examples. The hypotheses arising from these will give us a good starting point for searching for relevant

data in an electronic corpus. So, constructed examples have the advantage of being specially well suited for hypothesis generation. On the other hand, as is well known, many aspects of language can only be detected and tested against large scale corpus data.

In this article I will discuss Norwegian DNs with mainly constructed examples in LGP. This will give me a starting point for identifying relevant data in a corpus based investigation.

2. The Lexicalist Hypothesis

One of the most discussed questions in modern linguistic theory is the relationship between the syntactic and the lexical/morphological component of language. According to Chomsky's generative grammar the different linguistic levels are autonomous and well defined components or modules of grammar. This is often referred to as the **modular approach**. In most variants of functional grammar this relationship is rather seen as a gradient transition (cf. Andersen 2007b). The latter approach can be referred to as the **continuity approach**. The modular approach is compatible with Chomsky's lexicalist hypothesis. Chomsky (1970) argues that a theory of derivational morphology must be independent of a theory of syntactic transformations. The lexicalist hypothesis claims that syntactic rules cannot "see" structures below the level of the word. Consequently, nouns corresponding with verbs cannot be derived from these verbs, but have to be represented in lexicon as nouns. This hypothesis means that lexicon and syntax are separated into different modules of the language. These modules cannot be described with the same set of rules, i.e. they are autonomous and mutually exclusive systems. As will be demonstrated, some Norwegian DNs exhibit a parallelism between an unpacked syntactic phrase level and a packed morphological/lexical level.

3. Compound Deverbal Nouns

Intuitively the most natural assumption is that a DN is derived from a corresponding verb. But in many cases it is impossible to determine the direction of the derivation. In some cases the starting point seems to be the DN and the derived structure a corresponding verb (called back formations, cf. Faarlund et al. 1997:127): *åpne sesong* ("open season") → *sesongåpning* ("season opening") → *sesongåpne* ("to season open"). In this case the direct object of the verb construction seems to be incorporated (in the sense of Baker 1988) into the verb phrase before nominalization. In other cases the verb phrase does not have incorporation, like *bygge hytte* ("build cabin") → *hyttebygging* ("cabin building") → * *hyttebygge* ("to cabin build").

Based on compositionality Lyons (1977) makes a distinction between **compound lexemes** and **syntactic compounds**. The latter are compositional and regular. They have a very high type frequency and cannot be listed in the lexicon. However, the former group of compounds is non-compositional. Often they originate as syntactic compounds and develop specialized meanings. They become petrified and later completely fossilized. According to Lyons petrification and fossilization are two different stages of lexicalization of compounds. As shown in Andersen (2007a) there is evidence to indicate that the Norwegian DNs go through some of these phases in their historical development, Some examples will illustrate this general tendency of compounds. A noun like *appelsin* ("orange") was historically a compound consisting of apple+china. In present Norwegian this noun is not analyzed as a compound. It has been completely fossilized. In contrast a DN like *møbelsalg* ("furniture sale") is ambiguous. It may mean "current selling of furniture", in which case it is

compositional. But it may also mean "current selling at a cheaper price than the standard prize"), in which case it is non-compositional. The first meaning element (current selling) corresponds to the meaning contained in the corresponding verb, whereas the second meaning element (the cheaper price than standard) does not have a corresponding meaning inherited from the verb. This "extra" meaning element has developed in the DN only. The unpacked version will code this difference in the use of different prepositions. The first meaning can be unpacked as *salg **av** møbler*, the second meaning as *salg **på** møbler*. The idiosyncratic additional meaning element in the second meaning is an indication that the compound is in the process of being petrified.

4. Common Compound Types in Norwegian

In Standard Norwegian *bokmål* ("book language") there is a morphological distinction between compounds marked by the genitive enclitic marker –s-, as in *stat-s-kupp* ("coup d'état") and compounds with no markers, like *nett-Ø-bank* ("internet bank"). The distribution of these two types is largely unpredictable and lexically determined. The DNs behave like other compounds in this respect.

However, the semantic classification of Norwegian compounds show several relevant properties. In the most common type, determinative compounds, the final element is the head of the compound. The meaning of the first element delimits or modifies the meaning of the final element (typically a noun), like in *bil+hjul* ("car+wheel"). The first element, *bil*, tells us which type of wheel is meant (cf. Faarlund et al. 1997:66). Aikhenvald (2007:40) calls this type **endocentric compounds** to distinguish them from **exocentric compounds** like *dum+skalle* ("stupid"+"scull", i.e. a person who has a stupid scull, who is stupid). This latter type is also called possessive compounds and metonymy is a central part of the meaning construction.

5. The Direct Object Argument

A compound DN like *tre+felling* ("the cutting down of trees") cannot, however, be analyzed as a determinative compound, but it does correspond to the unpacked phrase *felling av trær* (with the same meaning) and the corresponding verb construction *felle trær* ("to cut down trees"). In the latter case the verb *felle* is a causative transitive and *trær* is the direct object. In other words, *felling* has the ability to allocate argument structure (AS), and the preposition *av* is a grammatical marker, a theta transmitter in the sense of Grimshaw (1990:70ff.).

A crucial question is then whether the incorporated element *tre* in *trefelling* can be shown to be a licensed argument (i.e. have a grammatical coding property for the transferred direct object function). If this is the case it would mean that organisation of AS, which is a typical syntactic phenomenon, is active also below the level of the word. This would run counter to Chomsky's lexicalist hypothesis.

In order to test this, we have to investigate whether AS organisation and licensing (i.e. grammatical coding) are operative below the level of the word. Recent investigation has showed that the inherent aspect of the verb is a crucial factor that interacts with AS licensing in DNs.

Most works dealing with aspect refer to Vendler's classical works (Vendler 1967, 1968). According to him there are four main verb classes: activities, states, accomplishments and achievements. The former two are atelic, i.e. they do not demand a climax and an ending, whereas the two latter are telic, demanding a climax and an ending. States are static as opposed to the other three dynamic classes. Accomplishments are imperfective in contrast to the perfective achievements. Vendler's own examples of states are *love*, *have* and *posess*. His examples of accomplishments are *draw a circle*, *paint a picture* and *make a chair*. Crucially, as can be seen from these examples, Vendler's classification is partly on phrase level and partly on lemma level. As I will demonstrate, the inherent aspect of the verb itself is often modified by the direct object argument in a verb construction, a phenomenon which on the lexical semantic level is called **co-composition** (Pustejovsky 1998:221ff.). Co-composition, then, means that the DN and its arguments combine to co-compose the meaning of the whole phrase or compound. If the result is compositional, the meaning of the constituent parts is the same as the meaning of the construction as a whole. The **inherent aspect** of a verb may be modified when arguments are added. The aspect of the phrase as a whole can be referred to as **co-compositional aspect**.

The inherent aspect of the verb *lese* ("to read") is that of an atelic activity in Vendlers terms. The naked corresponding DN *lesing* ("reading") has the same aspect classification. If you add the argument *en bok* (a book) the co-compositional aspect of the VP *lese en bok* changes to telic accomplishment. The corresponding DN construction *lesing av en bok* has again the same aspectual reading as the corresponding VP. The VP *lese bøker* and the corresponding DN *lesing av bøker* are both atelic activities, whereas *lese boken* and *lesing av boken* are both telic accomplishments. In some cases the aspectual meaning of the VP is underdetermined and has to be read off the context, as in [1] and]2]:

> [1] Å lese bok/boklesing er noe han ofte gjør.
>   To read book/the reading of books is something he often does
>   "He often reads books."

> [2] Å lese bok/boklesing denne ettermiddagen kunne han godt tenke seg.
>   To read book/the reading of books this afternoon could he well think himself
>   "To read a book on this afternoon was something he could well imagine."

The VP *å lese bok* ("to read book") with a naked generic object can be interpreted as an accomplishment, as in [2], or an activity, as in [1]. The same kind of underdetermination applies to the packed variant *boklesing*, which could be used instead of *å lese bok* in [1] and [2].

We can conclude from this that the use of the definite article (or determiner) in the postmodifying *av*-phrases changes the aspectual reading of the *–ing* DNs as regards telicity. These changes seem to correspond to the aspectual changes occurring with the corresponding VP constructions. This would indicate the fact that these DNs have argument structure in Norwegian. However, neither the grammatical preposition *av* (cf. Andersen 2007a), nor the lexical preposition *av* are coded in DN compounds. The DN *boklesing* is not only under-determined, but also polysemous in the above sense. In [1] and [2] the corresponding preposition *av* in the unpacked DN *lesing av bok* is a grammatical preposition, but *boklesing* can also be used when the first element *bok* is derived from a lexical use of *av* (as in [4] below):

[3] Forfatteren leste av boken sin.
   "The author read from his book."

[4] Etter lesing av boken sin, holdt forfatteren en forelesning.
   After reading from his book, the author gave a lecture.

[5] Etter boklesingen holdt forfatteren en forelesning.
   "After the book reading the author gave a lecture."

In other words, the lack of coding of the packed element in the compound DNs in [1], [2] and [5] would indicate that these constructions do not have argument structure in Norwegian. In order to say more about this unsolved question we have to analyse the basic functions of these participants. According to Grimshaw (1991) and Alsina (1996) the participant roles form a hierarchy where the roles lowest in the hierarchy are the ones most likely be incorporated into DN compounds as first elements, whereas the ones highest in the hierarchy are the ones least likely to be incorporated. Let us test Grimshaw's hierarchy. It looks like this:

[6] (Agent (Experiencer (Goal/Source/Location (Theme))))

Agent is always the most prominent participant and the most unlikely to be incorporated. Theme is the lowest participant and the most likely one to be incorporated. The Norwegian verb *overrekke* (hand over, give) have three participants, agent, goal and theme:

[7]

x	(y	(z))
Agent	Goal	Theme
Han overrekker	til folk	gaver
He gives	to people	gifts

We would then expect according to [6] that the theme would incorporate, but that the goal would not incorporate in the corresponding DN, and this prediction seems to hold:

[8] Gaveoverrekking til folk
   Gift-giving to people

[9]*Folkeoverrekking av gaver
   People-giving of gifts

[10] Overrekking av gaver til folk
   Giving of gifts to people

[10] shows that there is nothing wrong with *av gaver* in [9]. The unacceptability must be due to the first element *folk-* .

Similar other expressions in Norwegian seem to bear this out:

[11] blomsterdekorering i vaser/kakebaking for barn
   flower-decoration in vases/cake-baking for children

[12]*vasedekorering av blomster/*barnebaking av kaker
   Vase-decoration of flowers/children-baking of cakes

Some data from the Norwegian oil terminology show that Grimshaw's predictions are correct in some cases when applied to DNs in *–sjon* and *–ing*:

[13] *injisere vann* (Theme) *i brønn* (Loc)
     inject water   in well

[14] vanninjeksjon i brønn
     Water-injection in well

[15] *brønninjeksjon av vann

But, if only Loc is present, it can be incorporated, as in *brønninjeksjon*. Another example is *drenere vann fra havbunn* ("drain water from sea bottom") *drenering av vann* (Theme) *fra havbunn* (Loc) (the draining of water from sea bottom), *havbunnsdrenering*, *havbunnsdrenering av vann*.

I conclude that Grimshaw's prominence theory may contribute to the explanation of at least some elements of Norwegian DNs, both in LGP and in LSP. But this will preferably have to be investigated further on large corporate data. For many LSP variants this is not possible for Norwegian at the time of writing, but there exist national LGP newspaper corpora which are useful for a general investigation of these phenomena.

6. The Subject Argument

According to Grimshaw (1991) root DNs do not have argument structure in English. As I have pointed out in Andersen (2007a) this seems to apply to Norwegian root DNs with perfective or semelfactive aspect, as opposed to the imperfective process DNs in *-ing*. As [6] illustrates above, the agentive role is the least likely one to be incorporated in DN compounds. In the typical cases the agentive role is the subject role, referred to as the external argument by Williams (1981). As expected, the agentive role is unacceptable in imperfective process DNs, whereas the perfective/semelfactive roots allow this*: biestikk* ("bee sting") (perfective) vs *biestikking* ("bee stinging") (imperfective). In *biestikking* the first element is an external argument and is suppressed by a process called blocking (Sakshaug 1999:87ff.). In *biestikk* the first element is not the external syntactic argument (i.e. the subject) and hence not blocked, since this type of root nominals does not have AS.

Grimshaw also claims that the external argument must be saturated as the final one in compound DNs. This means that the external argument can only occur inside the DN if the other arguments are also present in the DN. In that case the head of the DN will be "saturated" and there is no open position. This is according to her not allowed. In Norwegian this hypothesis does not seem to be valid. In [16] both a causative/agentive and a non-causative/locative interpretation is possible:

[16] *legeundersøking* (doctor investigation, i.e. the fact that the doctor investigates, or: investigation at the doctor's), *hjerneblødning* (the brain is bleeding, or: there is bleeding in the brain).

This also applies to ergative verbs like in [17]:

[17] veiåpning (the road opens, or: someone opens the road), snøsmelting (snow melts or: someone melts snow), vektøkning (weight increases or someone increases weight), studentevaluering (students evaluate or someone evaluates students).

According to the blocking hypothesis these DNs should not have an inacusative, non causative interpretation. But other examples seems to contradict the blocking hypothesis unambiguously: *russefeiring* (the "russ" (i.e. high school students) are celebrating something) *russefeiring* can hardly be given an inacusative interpretation, i. e. that someone is celebrating the "russ". But unambiguous counterexamples like *russefeiring* are very few (cf. Faarlund et al. 1997).

In other DN compounds, however, the blocking hypothesis seem to predict correctly, like DNs in [18] (inacusative and hence blocked) and [19] (causative and hence not blocked), derived from the ergative verb *knuse* (break), unlike the DN derived from *smelte* in [20] and [21].

[18] *Glassknusing kan forårsakes av lydbølger.
    "The breaking of glass may be caused by sound waves."

[19] Glassknusing på offentlig sted er forbudt.
    "The braking of glass in public places is forbidden".

[20] Snøsmeltingen begynner sent i fjellet i år.
    "Snow melting starts late in the mountain this year".

[21] Snøsmelting går fortere med varmekabler.
    "Snow melting is quicker with heat cables".

In other cases it is difficult to identify the source of the agentive role in DNs. This is especially the case with DNs referring to artefacts. Artefacts are often contrasted with natural kind terms. According to Pustejovsky (1998:98) activities associated with artefacts, like *knives* and *cookies*, is the fact that they are made by human beings, in contrast to natural kind terms like *stones*, *trees* and *rivers*, where activities are typically associated with change of states without human intervention. Artefacts, then, have in common an agentive role which is an integral part of their lexical properties. This role has the same coding properties as the agentive role associated with AS, i.e. a premodifying genitive. The following examples will illustrate this:

[22] **Myndighetenes** bygging av sykehus.
    "The government's building of the hospital"

[23] **Ibsens** bok om Peer Gynt.
    "Ibsen's book about Peer Gynt"

[24]**Knuts** tre var hugget ned.
    "Knut's tree was chopped down."

[25] **Arkitektens** tegning av huset
    "The architect's drawing of the house"

Many DNs in most languages contain what Pustejovsky calls **logical polysemy** (1998:31). The typical type associated with DNs is the distinction between process and product. Some DNs, like *bygging* in [22] has only process meaning, whereas others, such as *tegning* in [25] has both a process meaning and a product meaning. The source of the premodifying agentive genitive in [22] is undoubtedly AS.

However, the role of the premodifying genitive which cannot be traced to AS is generally very vague. When the genitive premodifies an artefact, like in [23], the relation between the genitive (*Ibsen*) and its head (*book*) is very vague and underdetermined (cf. Andersen 2002). But in proper contexts an agentive role can be identified. In [23] it is the case that Ibsen wrote this book. *book*, being an artefact, has an agentive role which are activated in proper contexts. Pustejovsky refer to these as **qualia structures**. The source of the agentive cannot be AS, since *artefacts* such as *books*, do not have AS. Instead the source is to be found in the lexical semantics of these types of words.

In [24] the agentive role of the genitive is impossible, since the head denotes a natural kind term. The process meaning of *tegning* in [25] has AS, and the source of the genitive is AS, but if the product meaning is the one intended in [25], the product will have the status of an artefact. In this case an agentive interpretation would have its source in the lexical semantics of the artefacts.

In other words, DNs denoting artifacts are ambiguous as regards the source of the agentive role. If process meaning is intended, the source is AS, if product meaning is intended, the source is qualia structure. The first source is associated with syntactic phenomena, the second source is deeply situated in the lexicon. But, as I have demonstrated, AS seems to be active to some extent below word level. This runs counter to Chomsky's lexicalist hypothesis and to the modular approach, but gives further support to the continuity hypothesis.

7. Petrification, Fossilization and Terminologization

According to the continuity hypothesis syntactic constructions move from the syntactic level in the direction of the lexical level through a process called lexicalization (Brinton and Traugot 2005). This is also a strong tendency for the DNs in languages (Andersen 2007a). Petrification and fossilization can be considered as two stages in this diachronic process. These DNs move away from their corresponding verbs to become full fledged nouns, where the correspondence between the noun and the verb can no longer be detected. In this process the DNs become semantically enriched, i.e. new semantic features are added to the DN, features that are not present in the semantics of the corresponding verb.

An interesting question is whether this semantic enrichment can be specific to LSP domains. An initial hypothesis would be that domain specific semantic components, or rather characteristic features of terms, may be a result of common operative norms in a domain specific discourse community. As pointed out in Andersen (2007a:67f.), when DNs are coined productively in technical texts, the resultant DN is compositional, i.e. the verbal meaning inherent in the verb plus its participant roles are carried over in the nominalization process. The only additional element added in the process is the reference function inherited from the noun class property. This inheritance is dictated by the language system itself.

In addition to these compositional elements there is a potential for lexicalization-/terminologization, or in lexicalization terms "enrichment". In this diachronic process where syntactic compositional compounds move in the direction of idiosyncratic compound lexemes, the enrichment potential is realized. This is dictated by pragmatics, i.e. language use. As DN syntactic compounds are being used over time, language users tend to "invest" additional meanings into them. In an LGP context this is called lexicalization. If a similar process can be detected in various LSP variants, the process would be one of terminology-

zation. One way of investigating this hypothesis would be to compare comparable LSP texts from different time periods. This hypothesis is in harmony with the continuity approach where the transition between syntax and lexicon is gradual. The modular approach and Chomsky's lexicalist hypothesis only partly predicts correctly for Norwegian. To me this implies that the Norwegian system of DNs is in a state of flux and should be investigated further in a continuity approach.

References

Aikhenvald, Alexandra Y. (2007) Typological distinctions in the word formation. In: Shopen, Timothy (ed.) *Language typology and syntactic description*, Vol. 3: *Grammatical Categories and the Lexicon*. Cambridge: Cambridge University Press. 1-66.

Alsina, Alex (1996) *The role of Argument Structure in Grammar. Evidence from Romance*. Standford: CLSI Publications.

Andersen, Øivin (1998) *Lingvistikk og dokumentasjonsbaserte tesauruskonstruksjoner* [Linguistics and Documentary Based Thesaurus Constructions]. Bergen: HIT-senterets publikasjonsserie nr.1/98.

Andersen, Øivin (2002) Terminology and pragmatics. In: Koskela, Merja /Laurén, Christer /Nordman, Marianne /Pilke, Nina (eds.) *Porta Scientiae* I. Vasa. Vaasan yliopiston julkaisuja. 136-147.

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: Bassey, Edem Antia (ed.) *Indeterminacy in Terminology and LSP*. Amsterdam/Philadelphia. John Benjamins. 3-14.

Andersen, Øivin (2007c) Deverbal nouns as hybrids: Some textual aspects. In: Ahmad, Khurshid / Rogers, Margaret (eds.) *Evidence-based LSP. Translation, Text and Terminology.* Bern: Peter Lang. 155-171.

Baker, Mark. C. (1988) *Incorporation. A Theory of Grammatical Function Changing*. Chicago: The University of Chicago Press.

Brinton, Laurel J. /Traugott, Elizabeth Closs (2005) *Lexicalization and Language Change*. Cambridge: Cambridge University Press.

Chomsky, Noam (1970) Remarks on Nominalization. In: Jacobs, Roderick A. /Rosenbaum, Peter S. (eds.): *Readings in English Transformational Grammar*. Waltham, Mass.: Ginn and Co.

Faarlund, Jon Terje /Lie, Svein /Vannebo, Kjell Ivar (1997) *Norsk referansegrammatikk*. Oslo: Universitets-forlaget.

Grimshaw, Jane (1990) *Argument structure*. Cambridge, Mass.: MIT Press.

Jackendoff, Ray (1972) *Semantic Interpretation in Generative Grammar*. Cambridge: MIT Press.

Lyons, John. (1977) *Semantics* 2. Cambridge: Cambridge University Press.

Pustejovsky, James (1998) *The Generative Lexicon*. Cambridge: MIT Press.

Sakshaug, Laila (1999) *Norwegian Compound Deverbal Nouns: An Autolexical Analysis in Morphology, Syntax and Semantics*. Dr. artium thesis. Trondheim: Faculty of Arts.

Vendler, Zeno (1967) *Linguistics in Philosophy*. Ithaca, N.Y.: Cornell University Press.

Vendler, Zeno (1968) *Adjectives and Nominalizations*. The Hague: Mouton.

Williams, Edwin (1981): Argument Structure and Morphology. *Linguistic Review,* 1. 81-114.