Anders Søgaard
Center for Language Technology

The Grammaticalization and Disappearance of Adpositions in Nominal Compounds

Abstract

This article identifies a semantically motivated typology of lexicalization paths to account for the formation of nominal compounds from complex noun phrases. The typology includes coordination, semantic roles and possession, which, systematically, seems to be the origin of various subclasses of appositional, copulative and endocentric compounds, and the typology has serious implications for linguistic theory.

1 Introduction

This article tells the simple story of various kinds of complex noun phrases which over time have turned into binominal compounds. It identifies the paths of lexicalization and a number of important states occupied by unrelated languages across the world. The lexicalization of complex noun phrases is not itself an instance of grammaticalization, but the adpositions of those noun phrases exhibit every hallmark of that process as they are reduced from lexemes to morphemes and then finally dissapear. The major route which subsumes all the more specific and deterministic paths, can be roughly presented by the schema in (1):

(1) $[\alpha\otimes[]_{NP}]_{NP}\Rightarrow[\alpha\otimes[]_{N}]_{N}\Rightarrow[[]_{N}]_{N}$ where $\otimes$ is an unordered composition relation

The article presents many instances of this schema and identifies a number of more specific lexicalization paths. The α represents the adposition, or, in a "later" state, an affix. It is shown how adpositions of noun phrases are grammaticalized as compound markers and linking elements in the course of the lexicalization process, and how they often disappear by the end of this process.

The more specific paths constitute a fixed typology, it is argued. Interestingly, the paths are deterministic in the sense that the compound which is the result of the lexicalization process is largely determined by a corresponding noun phrase structure. In other words, the semantics of the adposition often identifies the constructional class to which the nominal compound ultimately belongs. Briefly, complex noun phrases which are constituted by prepositional and possessive phrases translate into endocentric compounds, while coordinated noun phrases translate into adpositional and copulative compounds. Some more specific subpaths are also identified. This may lead to important revisions of contemporary theories of the semantics and morphosyntax of nominal compounds.

1.1 Related literature

The literature on the semantics of nominal compounds is rich; see Søgaard (2005) for a review. The literature on the morphosyntax of nominal compounds is also relatively elaborate. A historical and typological view on matters is seldom adopted, however. This is somewhat surprising, since the motivation for such studies is quite obvious:

Lieber and others have noted a parallellism in head-modified ordering within NP's and compound nouns. Why should that be? Obviously, because compounds are lexicalised NP's. (Delancey, 1993)

Wälchli (2003) calls this rather intuitive hypothesis about the origin of compounds the *condensation hypothesis*. It says that patterns of compounds always derive diachronically from semantically corresponding syntactic constructions. This hypothesis is evidenced by much of the data presented here, but comes with a qualification. Once patterns of compounds are established, novel compounds may eventually lose any relation to the etymological origin of the construction. An important mechanism here is analogy (Ryder, 1994). Consider also the Danish data in the next section.

It is often hypothesized that languages search for equilibria, i.e., states which are economic and enable efficient communication. Such a functional requirement suggests some degree of conservation in language change, and predicts that language development or grammaticalization is best conceptualized as a transition from state to state. The hypothesis is not tested here, but different states are identified. Such states include juxtaposed compounds (English, Tagalog, etc.), compounds with affixation (Finish, Sanskrit, etc.), various intermediate states (Danish, German, etc.), and adpositional compounds (French, Italian, etc.). In the course of our investigations, more languages are associated with these states.

The last part of this section presents some illustratory examples of Danish compounds and their etymological history. The second section presents a brief typological survey of a more quantitative nature, though no actual statistics are presented. It rather presents the range of evidence for the condensation hypothesis and the constructions involved. The penultimate section discusses the implications of our investigations for linguistic theory. In particular, our results have consequences for the usual classification of compounds and for the debate about the morphosyntactic status of nominal compounds. I said that a historical and typological view on this debate was seldom adopted. This is exacly what is done here. This article is, in other words, a modest contribution to the sparse literature on the history and typology of nominal compounds.

1.2 A couple of examples from Danish

In the rest of this section which is only meant to briefly illustrate the phenomenon addressed below, the history of two Danish compounds is reviewed, namely *saftevand* and *barnsnød*:

(2)

saft	-e-	vand
juice	LINK	water

'diluted cordial'

(3)

barn	-s-	nød
child	LINK	pain

'birth pangs'

The linking element of (2) is a reduced conjunction (Bauer, 2001:699), now phonologically realized as a schwa. The *-e-* is a frequent linking element in Danish, but sometimes it is the trace of a genitive or a plural affix rather than a conjunction. In other words, a linking element does not reflect its own origin in direct ways. The second frequent linking element in Danish is *-s-*.[1] In (3), it is etymologically a genitive.[2]

It is important to note not only that the early adpositions syncretize, and linking elements thus conceal their past, but also that not all linking elements are traces of adpositions. Consider, for instance, (4):

(4)

år	-s-	tid
year	LINK	time

'season'

It seems reasonable to hypothesize that the *-s-* here, as in (3), was once a genitive. Unfortunately, or interestingly, the hypothesis is not a hundred percent true. The compound *årstid* is rather a direct translation of a German compound, *Jahreszeit*, which may of course have a genitive origin, but it is far from clear if this motivated the linking element of the Danish equivalent. Similarly, there is no clear motivation of the *-s-* in (5-7). (5) is a direct translation of the non-genitive English compound *gunboat diplomacy*. Compare (6) to the Swedish, non-genitive equivalent *bygdesemester*. (7) is a direct translation of the non-genitive English compound *credibility gap*; this can also be compared to the Swedish, non-genitive equivalent *förtroendeklyfta*. The Swedish data says a lot, since genitive compounds are as frequent in Swedish as they are in Danish, and since the two languages are similar in many respects.

(5)

kanonbåd	-s-	diplomati
gunboat	LINK	diplomacy

'gunboat diplomacy'

(6)

bondegård	-s-	ferie
farm	LINK	holiday

'holiday on a farm'

(7)

tillid	-s-	kløft
credibility	LINK	gap

'credibility gap'

If we zoom out, this is the picture we get: The linking elements, more generally, lost their semantics at some point in the history of Danish, and novel compounds now select their elements on different grounds. Only some compounds reflect a phrasal history.

The obvious reason for this assymmetry is of course lexicalization itself. Once a class of compounds is established as a lexical class, new compounds are produced. As it turns out, the linking elements are also lexicalized, and occur in novel compounds in predictable ways. The relevant factors in Danish seem to be in part phonological and in part lexical in nature, i.e. various lexical items go with certain linking elements, and certain phonological rules also govern the distribution of these elements.[3] Such rules are of no direct concern to us, but it is important for our investigation to distinguish between the lexicalization process itself, and the productive morphosyntactic patterns which are results of the lexicalization process. It is also important to note that the linking elements in Danish no longer come with a semantics. They are semantically empty elements, only realized by phonological rules. This is very different from, say, Italian, where linking elements still identify important compound classes.

In the next section, as mentioned above, a typological survey of similar lexicalization patterns is presented. It is shown how various compound classes etymologically reflect different kinds of complex noun phrases.

2 A typological survey

(2) and (3) gave evidence that both conjunctions and genitives can be reduced by way of condensation. The examples were from Danish, but of course this is not the only language on the globe which exbitits condensation effects. Take first the coordinative example in (2). Similar evidence is found in many other languages, from Ancient Indo-Iranian to Burushaski (Wälchli, 2003), but there is also some important variation to these examples. The compound in (8), for instance, is from Mari:

(8)

iz	-ak-	šol’	-ak-	šamyč
elder.brother	LINK	younger.brother	LINK	PL

‘brothers’

The double affixation indicates that this construction is not phrasal, and the *-ak* is not a conjunct, but an additive suffix. The condensation process not only reduces phonological form, though this is no doubt the tendency, but in Mari, it actually led to a kind of reduplication effect. Consider also an adjectival compound of Classic Greek in (9). The *-k-* is a reduced conjunction (*kaí*).

(9)

kalo	-k-	agathía
beautiful	LINK	good.ABST

‘beautiful and good’

It seems the condensation of coordination is not restricted to nominal conjuncts. The path which underlies the Classic Greek example, is represented in terms of our general schema, which was introduced in (1), but with a change of syntactic categories:

(10) $[kaí \otimes []_{AP}]_{AP} \Rightarrow [k \otimes []_{A}]_{A}$

(In the rest of the paper, only nominal compounds are considered. This example is just meant to illustrate the range of coordinative compounding.) (3), on the other hand, reflected an underlying genitive, or more accurately, it was coined from a genitival construction. Similar examples are widespread among the world's languages. Consider, for instance, (12) from German, (13) from Finnish, and (14) from Turkish. The Turkish compound does not reflect an adpositional construction, but the morphological suffixes indicate its origin as a (morphological) genitive. In less analytic languages, e.g. Russian, such patterns are of course more common. In our exposition, the patterns are represented by schemas only minimally different from (1), i.e., the affix attaches to a noun rather than a noun phrase:

(11) $[\alpha \otimes []_N]_{NP} \Rightarrow [\alpha \otimes []_N]_N \Rightarrow [[]_N]_N$

(12)

Kind	-s-	mutter
child	LINK	mother

'mother of a child'

(13)

auto	-n-	ikkuna
car	LINK	window

'car window'

(14)

otelin	odalari
hotel.GEN	room.PL.its

'rooms of a hotel'

Latvian has two compound constructions, one in which the modifying element occurs in its stem form, and another in which it occurs in the genitive. This illustrates the last steps of the envisaged path (17) nicely:

(15)

grāmat-veikals
book.store
'book store'

(16)

grāmat-u	veikals
book.GEN	store

'book store'

(17) $...\Rightarrow[u\otimes[]_N]_N\Rightarrow[[]_N]_N$

In Italian, as briefly mentioned, a great number of linking elements are employed. Originally, these were prepositions (and most of them still are, in the relevant contexts). The compound status of the examples presented here is evidenced by the fact that no adjuncts can intervene between the compound constituents.

(18)

coltello	di	ghiaccio
knife	LINK	ice

'knife made of ice'

(19)

Coltello	da	ghiaccio
Knife	LINK	ice

'knife for cutting ice'

Other prepositions which have similar uses, include *a* and *per*. The paths are of this form:

(20) $[per\otimes[]_{NP}]_{NP}\Rightarrow[per\otimes[]_N]_N$

More exotic linking elements are found in Maidu and Yimas (Bauer, 2001). In Maidu, the nominative marker is used as a linking element, while the oblique marker is used in Yimas. Sanskrit, much like Italian, presents a wider range of linking elements, incl. elements usually associated with accussative, instrumental, dative, locative and ablative morphology. The modifiers are apparently case-marked, and the cases identify compound classes (data and paraphrases gleaned from Bauer, 2001):

(21)

Dhana	-m-	jayá
Wealth	LINK.ACC	winning

'winning wealth'

(22)

Diví	cara
sky.LOC	moving

'moving in the sky'

(23)

vāc	-á-	stena
incantation	LINK.INST	stealing

'stealing by incantation'

The path for (23) is based on the affix schema in (11):

(24) $[\acute{a} \otimes []_N]_{NP} \Rightarrow [\acute{a} \otimes []_N]_N$

Similar data is found in Finnish. The linking element in (13) is etymologically a genitive, but all the local cases can be used as linking elements, which usually, much as in Turkish, attach to the modifying element. Consider, for instance (25), where the linking element reflects adessive morphology.

(25)

Matka	-lla-	oloaika
journey	LINK	be.DERIV.time

'the time the journey takes'

Such data is good evidence for the condensation hypothesis (Wälchli, 2003). Other evidence comes from a rather intimate correlation of word order in noun phrases and word order in nominal compounds. In Bauer's typological study, the word order of compounds seems to relate to the ordering of possessor and possessum phrases. (Out of 36 languages, there were only four exceptions to this correlation.) However, Bauer also found a number of inconsistencies. Some of these almost certainly relate to language change (Wälchli, 2003). It is not clear how etymologically induced distinctions affect this picture. The condensation hypothesis predicts there to be a correlate between word order of nominal compounds and word order of the related noun phrases. Since this relation may be rather complex, so can the word order distributions of course be complex.[4]

Some specific path schemas have already been abstracted. Our data suggests that such schemas come in various flavors, including paths whose sources relate to:

(a) coordination,

(b) semantic roles, and

(c) possession.

The paths can be shown to correspond to traditional compound classes, incl. (a) appositional and copulative compounds (or co-compounds in the terminology of Wälchli, 2003), and (b-c) endocentric compounds. The data presented in Søgaard (2005) give evidence that languages in fact grammaticalize the corresponding distinctions. In most contemporary European languages, with the exception of Modern Greek, for instance, appositional and copulative compounds are ungrammatical. The rest of this section is devoted to an explicate account of the abstract paths which license (a-c), and how they correlate with distributional facts about nominal compounds in natural languages. The paths are:

(26) $[\textbf{and}\otimes[]_{NP}]_{NP}\Rightarrow[\textbf{and}\otimes[]_{N}]_{N}\Rightarrow[[]_{N}]_{N}$

(27) $[\Sigma\otimes[]_{NP}]_{NP}\Rightarrow[\Sigma\otimes[]_{N}]_{N}\Rightarrow[[]_{N}]_{N}$

(28) $[\textbf{poss}\otimes[]_{NP}]_{NP}\Rightarrow[\textbf{poss}\otimes[]_{N}]_{N}\Rightarrow[[]_{N}]_{N}$

Each path has a number of subpaths or more specific paths, even at the non-language-specific level. Our accounts of these systems of subpaths rely on Wälchli (2003) for coordination, Søgaard (2005) for semantic roles and Søgaard (2006) for possession. The convention of Søgaard (2005) to use Σ as a variable for semantic roles is adopted. (One can, it should be noted, cross-classify the paths and the states and define the inventories of natural languages in a complete way.)

2.1 Coordination

Wälchli (2003) hypothesizes that certain kinds of tight coordination lead to compound formation over time. The notion of tight coordination, which he believes to be of typological

importance, is rather fuzzy though, and defined along several dimensions. Structurally, tight coordination is indicated iconically, e.g. by the length of the conjuncts, and morphosyntactically, by specialized markers. Tight coordination also has semantic correlates. Certain distinctions, found in the typological literature, of relevance here include:

(i) natural vs. accidental coordination,

(ii) group vs. separate coordination,

(iiii) intersective vs. non-intersective coordination, and

(iv) exhaustive vs. non-exhaustive listing coordination.

The distinction between natural and accidental coordination is well-studied. Some relevant literature includes Dalrymple and Nikolaeva (2005), who, for instance, point out that Udihe distinguishes between the two kinds of coordination at the surface level:

(29)

bi	Sergej	zu**n**e
I	Sergej	CONJ

I and Sergej'

(30)

Bi	mamasa	mule
I	wife	CONJ

I and Sergej'

The postposition *zu**n**e* is used for accidental coordination, and *mule* is used for natural coordination. Babungo, on the other hand, exemplifies a language which has grammaticalized the distinction between group and separate coordination (Wälchli, 2003), e.g.:

(31)

Làmbí	gh**ó**	Ndùlá	g`**e**	táa	yìwì**n**
Lambi	and	Ndula	go.PFV	to	market

Lambi and Ndula went to the market (together).'

(32)

Làmbí	n`**e**	Ndùlá	g`**e**	táa	yìwì**n**
Lambi	and	Ndula	go.PFV	to	market

'Lambi and Ndula went to the market (together or independently).'

The conjunction *gh***ó** is used only for group coordination. Similar phenomena are attested in Aymara. The distinction between intersective and non-intersective coordination corresponds to the distinction between appositional and copulative compounds. The distinction is set-theoretical. A compound such as *actor-producer* is, for instance, appositional. The more technical term, *space-time*, is copulative, on the other hand, since it refers to four-dimensional space rather than the intersection of space and time (whatever that might be). Is this distinction between intersective and non-intersective coordination reflected at the surface level of natural languages? Consider this data from English:

(33) That soldier and sailor were good friends.

(34) John's friend and colleague was here.

and their translation equivalents in Italian:

(35) *Quel soldato e marinaio erano buoni amici.

(36) L'amico e collaboratore di Gianni è stato qui.

It seems that for the construction $[_{D}[_{N}\text{ and }_{N}]]_{NP}$, no intersective reading is available in Italian. In (33), reference is made to two distinct individuals, evidenced by the verbal morphology. This split reading is not possible in Italian. Similar split readings are possible in a small set of languages, incl. Dutch and Finnish, but not, for instance, in French or Modern Greek. For a thorough investigation of this phenomenon, see Heycock and Zamparelli (2003).

The distinction between exhaustive and non-exhaustive listing coordination is grammaticalized in Kanuri, for instance. The associative plural and collective marker *-so* is used for non-exhaustive lists, while *-Cà*, which also functions as proprietive, is used for exhaustive lists.

Certain hypotheses can be formulated on the basis of this typology of subpaths, but they must be carefully evaluated. For instance, it seems rather intuitive to say that languages which do not allow split readings in tight coordination, should neither allow copulative compounds. However, Modern Greek provides a counterexample to this hypothesis in that it blocks split readings, but exhibits copulative compounds.

Wälchli (2003) provides strong evidence for a natural transition from natural coordination into copulative compounds. Søgaard (2005) claims the important constraints on copulative compounds are taxonomic of nature. Does such a constraint apply to natural coordination as well, as would then be predicted by the condensation hypothesis? Consider the natural pairs in the data set of Dalrymple and Nikolaeva (2005). They consist of kinship terms (*father* and *mother*, *son* and *daughter*, etc.) and rather idiosyncratic conceptual pairs, incl. *sun* and *moon*, *salt* and *bread*, etc. Every pair in the data set obeyed the taxonomy requirement of Søgaard (2005), i.e. they were of equal taxonomic rank.

In Danish, as in many other European languages, tight coordination is restricted to natural coordination. Remember the copulative compound in (2). On the basis of our Danish data alone, a hypothesis can be formulated which is parallel to the intuitions of Wälchli (2003). Schematically, the hypothesis can be represented as:

(37) $[og\otimes[]_{NP}]_{NP} \Rightarrow_n [og\otimes[]_N]_N \Rightarrow [e\otimes[]_N]_N \Rightarrow [[]_N]_N$

The transition $\Rightarrow_n$ requires the conjuncts to obey the constraints of natural coordination. The rest of the path determines the transition of the natural coordination to a link-free compound. The condensation path is not (yet) complete for Danish. As mentioned in the introduction, Danish is in an "intermediate" state with regard to the relevant processes. The conservativity imposed by equilibria may of course slow or even hinder this development.

The (informal) semantics of this construction is presented in (38). The constituents are δβ, [[δ]] is the denotation of δ, [[δ↑]] is the denotation of the superconcept of the semantic concept

that corresponds to δ, and ⊗ is the mereological operator that is a function from entities to fusioned entities, i.e. conceptual pairs. The superscript τ indicates raising of the denotation to kinds or types (intensional readings).

(38) $\left\{ \begin{array}{l} [[\delta]] \wedge [[\beta]] \\ [[(\delta) \uparrow \vee (\beta) \uparrow]]^{\tau} \\ [[\delta]]^{\tau} \otimes [[\beta]]^{\tau} \end{array} \right\}$

The first disjunct corresponds to the appositional reading, whereas the second and third correspond to the copulative. The appositional reading is intersective, the second reading denotes the superconcept of the concepts denoted by the constituents, whereas the third denotes a fusion. The second reading assumes that the constituents are mapped onto subconcepts of some other concept. See Søgaard (2005) for other determining factors, i.e. parsing techniques for compound interpretation.

2.2 Semantic roles

The Finnish and Sanskrit compounds presented above show how compound classes are often defined in terms of semantic roles, and how the distinctions reflect in surface morphosyntax. The analysis of Italian compounds presented in Johnston and Busa (1999) also refers to semantic roles. In this section, evidence is presented that natural languages in fact encode compound classes in terms of semantic roles. This is not clear from the Finnish and Sanskrit data. The fact that modifying elements exhibit case morphology, does not necessarily tell us anything about the encoding of lexical classes. If instead languages were identified in which certain combinations of semantic roles lead to ungrammatical compounds, preferably irrespectively of case-marking, this would count as harder evidence. This is exactly the kind of evidence presented here. The relevant data comes from Danish and English.

Compared to Danish and German, English compounding is more restricted. Specifically, it is almost impossible to coin an eventive compound in which the first constituent occupies an agentive or objective role (if the head noun is non-derived), e.g.:

(39) *smith hammer

(40) *child doctor

Some authors have proposed that this asymmetry is the result of a constraint that compound modifiers cannot be human in English. This account is rather problematic, however, since human modifiers are fine, once, for instance, the compound gets a metaphorical reading, e.g. *child bed*. Other counter-examples include *baker joke* and *doctor television*. On the other hand, the classification in terms of semantic roles also has a few counter-examples, incl. *dog food*. The question is undoubtedly more complex than what is suggested here, and maybe some cross-classification is needed. It is interesting, though, to compare the data to Danish or German equivalents, specifically:

(41) hammer smith

(42)

smedehammer
smith.hammer
'smith hammer'

(43)

*hammersmed
hammer.smith
'hammer smith'

The point is that in Danish and German, eventive compounds with instrumental modifiers are ungrammatical.[5] Are there any syntactic, phrasal correlates to these assymmetries of English and Danish and German? Apparently not. In addition, this is complicated by the fact that instrumental compound modifiers apparently were grammatical in Old Danish, e.g. (43) was both a surname and the name of an occupation.

It should in fact come as no surprise that there are no phrasal correlates to these asymmetries in the distribution of compounds, since the phrasal origin of endocentric compounds is in the

fully productive adpositional domain. The closest we get is probably certain pragmatic preferences, for instance:

(44) a teacher at the blackboard

(45) ? a computer at the blackboard

(46) a blackboard teacher

(47) ? a blackboard computer

The preposition *at* seems to require there to be some intimate (ontological) relation between its complement and the head noun. Such a semantic adjacency is often presupposed in compounds too. Consequently, the oddness of (45) is also reflected in (47). Apparently, "soft" pragmatic constraints on acceptability harden or lexicalize in compound classes. It is also possible that the constraints are results of type coercion (from token into types).

The (informal) semantics is introduced in (48). The new thing is that a dependent type relation Δ is introduced. δ is said to be the head noun, and *i*,*j* are indeces.

(48) $[[\delta]]_i^\tau \wedge [[\beta]]_j^\tau \wedge \Delta_\delta(i,j)$

This means that there is a relation, dependent on the head noun, that captures the relational semantics of the compound construction. The Δ relation may be decomposed in a neo-Davidsonian manner, and Θ-roles introduced, but intuitively, the semantics of, say, (18) and (19) could be reconstructed as[6]

(49) $[[coltello]]_i^\tau \wedge [[ghiaccio]]_j^\tau \wedge make\_of(i,j)$

(50) $[[coltello]]_i^\tau \wedge [[ghiaccio]]_j^\tau \wedge cut(i,j)$

2.3 Possession

The semantics of possession is a complicated domain, and we only briefly summarize the results of the typological survey in Søgaard (2006). This survey investigated the semantics of possessive constructions in English, German, Hebrew, Hocak, Italian, Japanese, Norwegian Bokmål, Russian and Yucatec Maya. It was shown how each language provided a set of

possessive constructions which covered the semantic domain of kinship, part-whole relations, qualia, and ownership. It was also evident from the data that these four subdomains constitute a linear hierarchy, in that particular order, such that a construction is always associated with a continuous span of this hierarchy. In Hocak, for instance, three possessive constructions were identified. One covers kinship, another covers part-whole relations. The third one, the juxtaposed possessive, covers qualia and ownership. Russian has two general constructions which cover the entire span of the hierarchy, and two specialized constructions, the prenominal genitive and the nominative pronominal possessive, which are only associated with ownership. In Yucatec Maya, the domain is partitioned by two constructions.

It would be interesting to see if a possible transition from possessives to compounds, suggested for instance in Bauer (2001), is limited to a certain span of this hierarchy. A very plausible suggestion, on pragmatic grounds, is that only possessives associated with part-whole relations, e.g. (51), and qualia, e.g. (3), undergo condensation, since such possessives often refer to types rather than tokens.[7] This seems to be a relevant criteria for condensation. In Western Indonesian languages, the type-token distinction is encoded by two separate possessive constructions. In these languages, the condensation hypothesis is also open for evaluation. The compound (51), which is included here for illustration, has a genitive linking element (Bauer, 2001:703):

(51)

jul	-e-	dag
Christmas	LINK	day

'Christmas day'

2.4 Grammaticalization beyond compounding

The story does not stop with compounds, though. This article is primarily concerned with transitions from phrases to compounds, but in fact there is evidence that compounds often fuse, a process sometimes referred to as syntactic coalescence. The story, briefly, is that compound

heads or modifiers become clitics, then affixes, and finally dissapear. Consider for instance, the etymology of diminutives. Often diminutive affixes come from nouns such as *child*. This, for instance, is the case in Classical Tibetan, Ewe, Fuzhou, and Thai. In Mandarin, the diminutive *-er* is etymologically a noun meaning 'son'. Jurafsky (1996) reviews the relevant data. The envisaged transitions, from phrase to diminutive, are illustrated by this in part fictitious example:

(52) woman like a child π child-like woman π child woman π child-woman or woman-child

The different constructions are all supposed to mean 'girl'. The first state is fully phrasal, the next is a compound with an adpositional marking, which dissapears in the next transition, and finally, the diminutive is coined. Such patterns seem to be quite realistic.

3 Implications for linguistic theory

It is common to identify four compound classes: appositional, copulative, endocentric and exocentric. The data presented here suggests that there are syntactic correlates of appositional, copulative and endocentric compounds, but Bauer (2001:700) notes that for languages which exhibit exocentric compounds "there is no apparent correlation with any structural facet". It is clear from his investigations, though, that languages with exocentric compounds all have endocentric compounds too, and it has been suggested in recent literature that exocentric compound classes are analogical extensions of endocentric ones (Ryder, 1994).[8] In this section, it is seen that the mere recognition of the effects of condensation has important consequences for contemporary theories of nominal compounds. The theories which relate to morphosyntax and those which relate to semantics, are treated separately.

It is a well-known controversy in the linguistics literature whether compounds are morphological or syntactic in nature. It is not evident that there is a universal answer to this question. Radford (1988) proposes an X' syntactic analysis, on which certain compound noun modifiers are complements, and others are attributes, e.g., Figure 1.

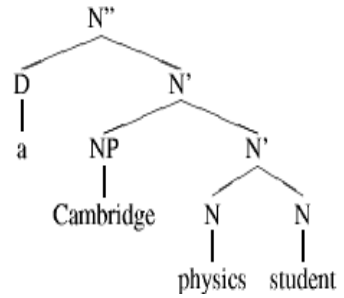

Figure 1: X' analysis of nominal compounds (Radford, 1988).

The evidence for such an analysis comes from the ungrammaticality of (53) and *one*-pronominalization.

(53) *a physics Cambridge student

In Danish, for instance, matters are slightly different. The kind of evidence obtained from *one*-pronominalization does not apply to Danish for several reasons. There is no good, non-archaic equivalent of pronominal *one*, for one thing, and since compounds constitute ortographic and phonological units, the data is very different. However, one fact remains. The complement must be adjacent to the head noun in a trinominal construction (the dashes are only ortograhically motivated):

(54)

En	Cambridge-fysikstuderende
A	Cambridge.physics.student

'a Cambridge physics student'

(55)

*en	fysik-Cambridge-studerende
a	physics.Cambridge.student

'a Cambridge physics student'

This article is only concerned with attributive compound modifiers, not complements. Consequently, the main point here is that compound modifiers apply recursively. Can this thesis be supported or qualified by our investigations? Intuitively, all the phrasal constructions that we have considered to be sources of compound constructions, are recursive, and thus this property already seems to be ensured by the condensation hypothesis. However, there is at least

one important qualification to be made on the basis of our investigations. Since only certain possessive constructions are recursive, this seems to limit the compound sources in the possessive domain. In the above, it was seen that only a subset of possessive constructions semantically qualify as sources of condensation, but in addition, this suggests that there are certain structural constraints to this path too. For instance, it predicts that the compound *Kindsmutter* in (12) has its origin in the recursive genitive (Figure 2), which actually seems to be the case.

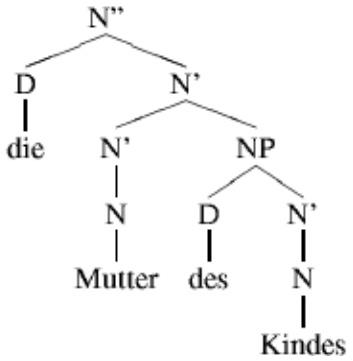

Figure 2: The German recursive genitive.

Contemporary theories of the semantics of nominal compounds come in four flavors (Søgaard, 2005). **Reductionist theories** claim that there is a fixed set of semantic relations, cognitively encoded, not linguistically, which determines the possible range of interpretations of nominal compounds. Several arguments against such theories have been proposed in the literature, e.g. that any such classification is arbitrary or incomplete. It seems that the condensation hypothesis provides an additional argument against reductionist theories, namely that since the semantic domain of phrasal constructions is open-ended, so is the set of compound classes which can be coined on this domain. The second set of theories, **transformational theories**, derive the meaning variation of nominal compounds from underlying structure. Of course the condensation hypothesis in part favors such an approach. The historical evidence runs counter the actual theories, however, most of which suggest that compounds are derived from relative clauses. The condensation hypothesis seems to provide

no evidence against or for **slot-filler theories**. Briefly, a slot-filler theory hypothesizes that constituents are conceptualized as bundles of semantic features, and the modifying constituent adds a feature to or specifies a feature in the semantics of the head constituent. Slot-filler theories have been faced with a number of challenges (Søgaard, 2005), incl. the productivity of exocentric compounding in some languages, but the condensation hypothesis does not help us to evaluate such theories. Finally, **pragmatic theories** claim that compounds are semantically unspecified, and that the meaning is resolved in context by pragmatic, knowledge-based inferences. The data presented in this article is clear evidence against such theories. Additional counterevidence is presented in Søgaard (2005).

4 Conclusions

In this short article, some data were presented in favor of the condensation hypothesis, i.e. that compounds derive diachronically from phrasal constructions. It was shown how condensation also interacts with the grammaticalization of adpositions. It was postulated that there is a typology of subpaths (and a cross-classification of paths and states) that govern condensation, and which relate sets of recursive phrasal structures with sets of recursive compound constructions. The sets are constituted by three major subsets, which were associated with coordination, semantic roles and possession.

Endnotes

[1]Only two other linking elements exist in Danish (Allan et al., 1995). None of them seem to be fully productive. They are *-en-*, which is borrowed from German and found in *bakkenbart* and *galgenhumor*, and *-er-* as in *blomsterbuket* and *nattergal*.
[2]The etymology of the Danish compounds in this section is gleaned from Juul-Jensen et al. (1919), if no explicit reference is given.
[3]Allan et al. (1995) suggest that juxtaposed compounds are licensed when the modifier ends in *-s*, a stressed vowel or unstressed *-e*, unstressed *-el*, *-en*, *-er* and when the ending is Romance, except *-ion*, *-tet* and *-um*. If the modifier ends in one of these tree suffixes, the result is an *-s-* compound. Such compounds are also licensed if the modifier ends in *-dom*, *-else*, *-hed*, *-ing*, *-sel* and *-skab*, or if the modifier is itself a compound. The *-e-* compound is licensed when the modifier ends in a consonant, and the head noun begins in one, if the modifier ends in *-ing*. Note that *-ing* is mentioned twice.
[4]An additional complication is language change. For instance, Italian and Vietnamese have both left-headed and right-headed nominal compounds. The patterns arose from diverse influences at different points in the history of the languages.
[5]Grammaticality judgments differ somewhat. Of course it is possible to get metaphoric readings of such compounds, e.g. *guitarpædagog*. Some people also seem to accept literal compounds such as *pistolrøver* (lit. 'gun robber').
[6]Modalities are also ignored in these simple semantic forms.
[7]A necessary qualification is that some compounds may be coined on kinship or possession relations, but then coerced into a type reading, e.g. (i) and (ii). The compound in (ii) may also be interpreted as a qualia compound, or it is licensed by some kind of bare noun modification.

(i) barn -e- barn
child LINK child
'grand-child'

(ii) læge-bil
doctor.car
'doctor's car'

[8]Søgaard (2005) argues against this view. It seems that certain kinds of exocentric compounds have no endocentric correlates. If so, the story becomes somewhat more complex and is likely to involve cognitive operations such as metonomy and metaphor in a more systematic manner. This is not of great concern to us.

References

Allan, R., P. Holmes and T. Lundskær-Nielsen. 1995. *Danish*. London: Routledge.

Bauer, L. 2001. Compounding. In M. Haspelmath et al. (eds.), *Language typology and language universals*. Berlin: Walter de Gruyter. Pp. 695–708.

Dalrymple, M. and I. Nikolaeva. 2005. Syntax of natural and accidental coordination. Unpublished manuscript, University of Oxford, England.

Delancey, S. 1993. Grammaticalization and linguistic theory. In J. de Garcia & D. Rood (eds.), *Proceedings of the 1993 Mid-America Linguistics Conference*, Boulder, Colorado, pp. 1–22.

Heycock, C. and R. Zamparelli. 2003. Friends and colleagues: plurality, coordination, and the structure of the DP. Unpublished manuscript, University of Edinburgh, Scotland.

Johnston, M. and F. Busa. 1999. Qualia structure and the compositional interpretation of compounds. In E. Viegas (ed.), *Breadth and depth of semantic lexicons*. Dordrecht: Kluwer.

Jurafsky, D. 1996. Universal tendencies in the semantics of the diminutive. *Language* 72: 533–578.

Juul-Jensen, H. (ed.). 1919. *Ordbog over det danske sprog*. Copenhagen: Gyldendalske Boghandel.

Radford, A. 1988. *Transformational grammar*. Cambridge: Cambridge University Press.

Ryder, M. E. 1994. *Ordered chaos: the interpretation of English noun-noun compounds*. Berkeley: University of California Press.

Søgaard, A. 2005. Compounding theories and linguistic diversity. In Z. Frajzyngier, A. Hodges & D. Rood (eds.), *Linguistic diversity and language theories*. Amsterdam: John Benjamins. Pp. 319–337.

Søgaard, A. 2006. The semantics of possession in natural language and knowledge representation. *Journal of Universal Language* 6(2): 85–115.

Wälchli, B. 2003. *Co-compounds and natural coordination*. Stockholm: Stockholm

Anders Søgaard
Center for Language Technology
Njalsgade 80, DK-2300 Copenhagen, Denmark
E-mail: anders@cst.dk