

COMPOUND CONSTRUCTION: SCHEMAS OR ANALOGY? A CONSTRUCTION MORPHOLOGY PERSPECTIVE

To appear in Sergio Scalise and Irene Vogel (eds.). *Compounding*. Amsterdam / Philadelphia: Benjamins

Suggested short title: Compound construction: schema or analogy

Author:

Geert Booij
Leiden University Centre of Linguistics
Faculty of Humanities, University of Leiden
P.O. Box 9515
2300 RA Leiden, The Netherlands
e-mail: g.e.booij@hum.leidenuniv.nl

April 2009

Abstract

This chapter argues that there is no absolute boundary between analogy and abstract schemas in word formation. Patterns of compounding are captured by constructional schemas of various degrees of abstraction. The necessity of such subschemas is argued for on the basis of observations on semantic specialization, headedness variation, diachrony, and allomorphy selection. Analogy and abstract schemas are opposite endpoints on a scale of schematicity.

Keywords : allomorphy, analogy, constructional idioms, Construction Morphology, derivation, headedness, incorporation, polysemy, subschemas

Word count: 6.560

COMPOUND CONSTRUCTION: SCHEMAS OR ANALOGY? A CONSTRUCTION MORPHOLOGY PERSPECTIVE

GEERT BOOIJ

University of Leiden

This chapter argues that there is no absolute boundary between analogy and abstract schemas in word formation. Patterns of compounding are captured by constructional schemas of various degrees of abstraction. The necessity of such subschemas is argued for on the basis of observations on semantic specialization, headedness variation, diachrony, and allomorphy selection. Analogy and abstract schemas are opposite endpoints on a scale of schematicity.

1. *Introduction: rule or analogy?*

Compounding is the best type of evidence for the claim that word formation should at least partially be accounted for in terms of abstract symbolic rules or schemas. In many languages the formation of new compounds is by far the most productive type of word formation. The formation of new compounds is not necessarily based on the model of existing compounds. Hence, compounding is often used to illustrate rule-governed creativity in the domain of word formation, and seems to be the best case for the theoretical position that word formation cannot be fully accounted for in terms of analogy.¹

It is obvious that analogical word formation does exist, as illustrated here by some examples from Dutch:

- | | | |
|-----|---|---|
| (1) | moeder-taal ‘lit. mother language, native language’ | vader-taal ‘lit. father-language, native language of father’ |
| | hand-vaardig ‘lit. hand-able, with manual skills’ | muis-vaardig ‘lit. mouse-able, with mouse-handling skills’ |
| | nieuw-komer ‘lit. new-comer, recent immigrant’ | oud-komer ‘lit. old-comer, immigrant who has arrived a long time ago’ |

For these words we can indeed point to one particular compound as the model word for the formation of the new compound, and the meaning of this new compound is not retrievable without knowing the (idiomatic) meaning of the model compound.

Analogical word formation may develop into a pattern that abstracts from specific model words. In English, the word *Watergate* functioned as the model of a number of English compounds in *-gate* that all denote a particular political or personal scandal, and hence this looks like a clear case of analogical word formation. However, since a set of such words has been formed in the meantime, it is no longer obvious that it is always the word *Watergate* itself that functioned as the model word. Once a set of words in *-gate* has been formed, language users may discover the commonality of such words in *-gate*, and hence this kind of productive compound formation is now better characterized by the schema:

$$(2) \quad [[x]_{Ni} [gate]_{Nj}]_{Nj} \leftrightarrow [\text{political scandal pertaining to SEM}_{i}]_j$$

(where SEM stands for the meaning of the co-indexed word constituent). That is, the word *gate* has acquired the meaning ‘political scandal’ when embedded in compounds. The assumption of such a schema expresses that it is no longer necessarily the case that language users model their new *gate*-compounds after the word *Watergate*. Dutch speakers have extended this use of *-gate* to Dutch, as illustrated by the following examples (Hüning, 2000):

- (3) kippen-gate ‘chicken-gate, scandal concerning chickens’
Stadion-gate ‘financial problems concerning renovation of the Olympic Stadium’
Zuid-Holland-gate ‘financial scandal concerning the province of Zuid-Holland’

Hüning concluded that the set of *-gate* words gave rise to a new morphological process in Dutch. This use of *-gate* is comparable to that of endings like *-burger*, *-holic*, *-tainment*, and *-zine* in English: a new type of compound constituent with a specific meaning is created, with the structural reinterpretation of an existing complex word being the first step. In the case of *hamburger*, the rise of *-burger* presupposed a morphological reanalysis: *hamburg-er* > *hamburger*, whereas in the case of *-gate*, the compound structure of *Water-gate* is maintained, but with a different interpretation of *gate*. A similar example from Italian is the emergence of *-poli* ‘scandal’ (Ramat, 2001). This use of *-poli* emerged from the word *tangento-poli* ‘rake-of town, town where the rake-of system is dominant’. New coinages are *sanito-poli* ‘the affair of the health ministry’ and *banco-poli* ‘bank scandal’.

In this chapter I will provide evidence that language users are able to discover compound schemas of various degrees of abstraction. At one end of the scale of abstractness we find analogical word formation based on a concrete model word, as illustrated above in (1). At the other end of the scale of abstractness, we have the following schema for Dutch compounds (and those in other Germanic languages) which expresses the generalization that Dutch compounds are right-headed:

$$(4) \quad [X_i Y_j]_{Y_k} \leftrightarrow [\text{SEM}_j \text{ with some relation } R \text{ to } \text{SEM}_i]_k$$

(where X and Y are variables for lexical categories). Between these two extremes, there are many generalizations about subsets of compounds that need to be expressed as part of the linguistic knowledge of the language user. This knowledge can be modeled in a hierarchical lexicon (Booij, 2005, Booij, 2007). In a hierarchical lexicon the set of established words is listed together with generalizations over subsets of words that share certain properties. Sets of words that share a particular form and corresponding meaning form the basis for discovering morphological regularities. In this chapter, arguments for such a model of the lexicon are provided by four different kinds of phenomena: the bound meanings of compound constituents (section 2), headedness variation (section 3), the emergence of derivation out of compounding (section 4), and allomorphy patterns (section 5) In section 6, I summarize my findings, and their implications for analogical models of morphology.

2. *Semantic subpatterns in compounding*

It is well known that many, if not all, derivational affixes derive historically from lexemes used as the first or second constituent of compounds. “Diachronically, the transmutation of a “blurred” compound into an affixal derivative is an almost trivial phenomenon” (Malkiel, 1978: 128). Lots of examples have been discussed in the literature on Germanic languages like Dutch, English, German, and Swedish (Ascoop, 2005, Ascoop and Leuschner, 2006, Booij, 2005, Booij, 2007, Dalton-Puffer and Plag, 2000, Leuschner and Decroos, 2008). The phenomenon is not so trivial as Malkiel suggested, since it reveals how the lexicon is organized: compound words beginning with or ending in the same constituent may form word families that can be characterized in terms of schemas for complex words in which one of the constituents is lexically specified.² When such a specified constituent lost its status as independent word, it could become an affix since it

survived as part of a compound schema, with sometimes less lexical and semantically more abstract meanings (Booij, 2005, Booij, 2007).

In other cases, the compound constituent still corresponds to a word that also occurs independently, but has acquired a specific meaning when part of a compound. In that case, one might speak of affixoids, but this is just a convenient descriptive term, not a new morphological category. In some cases, the meaning of the lexeme becomes more abstract, and may lose lexical content. An example of this phenomenon is the set of words with intensifying meaning that occur as the left constituents of Dutch XA compounds. Some examples are listed in (5).

(5)	<i>noun:</i>	<i>example</i>
	ber-e ‘bear’	bere-sterk ‘very strong’, bere-aardig ‘very kind’
	bloed ‘blood’	bloed-serieus ‘very serious’, bloed-link ‘very risky’
	dood ‘death’	dood-eng ‘very scary’, dood-gewoon ‘very ordinary’
	poep ‘shit’	poep-heet ‘very hot’, poep-lekker ‘very pleasant’
	ret-e ‘ass’	rete-leuk ‘very nice’, rete-spannend ‘very exciting’
	reuz-e ‘giant’	reuze-leuk ‘very nice’, reuze-tof ‘very good’
	<i>adjective</i>	
	dol ‘mad’	dol-blij ‘very happy’, dol-gelukkig ‘very happy’
	stom ‘stupid’	stom-toevallig ‘completely coincidental’, stom-verbaasd ‘very surprised’
	<i>verb</i>	
	kots ‘vomit’	kots-misselijk ‘very sick’, kots-beu ‘very tired of’
	snoei ‘prune’	snoei-hard ‘very hard’, snoei-heet ‘very hot’

These are clear cases of semantic reanalysis: the first constituent is reanalyzed as a morpheme with intensifier meaning. This reanalysis is made overt by the fact that these morphemes are attached to new semantic types of words. A noteworthy point concerning *bere-*, *rete-*, and *reuze-* is that they consist of a consonant-final stem followed by a linking element *-e* [ə]. This linking element is a necessary part of these nouns when used as intensifier prefixoids.

We can represent the affixoid nature of these compound-initial lexemes by specifying them in constructional idioms of the following form:

- (6) [[bere]_N [x]_{Ai}]_{Aj} ↔ [very SEM_i]_j
 [[dol]_A [x]_{Ai}]_{Aj} ↔ [very SEM_i]_j

[[loei]_V[x]_{Ai}]_{Aj} ↔ [very SEM_i]_j

[[wereld]_N[x]_{Ni}]_{Nj} ↔ [excellent SEM_i]_j

Constructional idioms are morphological or syntactic schemas in which one or more positions are lexically fixed, whereas other positions are open slots, represented by variables (Jackendoff, 2002). Being embedded in constructional schemas makes these words similar to prefixes. The only difference is that prefixes do not carry a lexical category label, and hence cannot be related to independent lexemes in the lexicon.

The specific meaning of intensification as illustrated in (5) is a precondition for these affixoids to be used in repetitive coordination that carries an emphatic meaning. This appears to be a systematic option for all prefixoids with intensifier meaning, as a Google search (13 May 2008) reveals:

- (7) a. bere- en bere-goed ‘very, very good’
bloed- en bloed-mooi ‘very, very beautiful’
dood- en dood-ziek ‘very, very ill’
poep- en poepheet ‘very, very hot’
rete- en rete-stabiel ‘very, very stable’
reuze- en reuze-tevreden ‘very, very pleased’
- b. dol- en dol-komisch ‘very, very comical’
stom- en stom-dronken ‘very, very drunken’
- c. kots- en kots-beu ‘very, very tired of’
snoei- en snoei-lelijk ‘very, very ugly’

The same kind of repetitive coordination is possible with Dutch intensifying prefixes such as *door-* and *in-* :

- (8) a. door- en door-nat
through and through-wet
‘wet through’
- b. in- en in-triest
in- and in-sad
‘very, very sad’

If we want to make a generalization as to which elements can occur in such repetitive coordination, we need to be able to refer to the class of compound-initial words with intensifier meaning. This is possible thanks to schemas like those in (6).³

The reality of the generalizations expressed in (6) is confirmed by the observation that in Dutch some of these prefixoids have developed into adjectives. This is the case for *reuze* and *kut*:

- (9) [reuze-]_N ‘giant-’ > reuze (A) ‘fantastic’
 [kut-]_N ‘cunt-’ > kut (A) ‘bad, worthless’

The following sentences illustrate this adjectival use of these two words:

- (10) a. Ik vind dat helemaal kut
 I find that completely cunt
 ‘I consider this completely worthless’
 b. Het uitstapje was reuze
 The outing was giant
 ‘The outing has been great’

The nominal origin of *reuze* ‘fantastic’ is reflected by its final schwa which is a linking element, the noun itself being *reus*. Such a development can only be understood if we assume a subpattern [[*reuze*]_NA]_A in the hierarchical lexicon of Dutch. The meaning of intensification of these nouns is a type of meaning expressed prototypically by adjectives, and hence the categorial reinterpretation of these nouns as adjectives in this context is a natural development.

Similar productive patterns have been observed for German and Swedish. In Swedish, for instance, we find *skit-bra* ‘shit-good, very good’ and *jätte-vinst* ‘giant-profit, very high profit’. The word *jätte* can also be used as an adjective, parallel to the Dutch word *reuze*. German examples are *Klasse-* in *Klasse-film* ‘class-film’ and *Spitzen-I* in *Spitzen-film* ‘top-film’ both meaning ‘excellent film’. (Ascoop, 2005, Ascoop and Leuschner, 2006).

An example of semantic concentration, a specific type of semantic development within compounds, is the use of the verb *scharrel* ‘to scratch. in the word *scharrel-kip* ‘lit. scratch chicken, free range chicken’. This words refers to chickens that can freely scratch the ground, and potter around. This use has been extended to other compounds:

- (11) a. scharrel-vlees

- scratch-meat
'free range meat'
- b. scharrel-ei
scratch-egg
'free range egg'
- c. scharrel-melk
scratch-milk
'free range milk'
- d. scharrel-wijn
scratch-wine
'eco-wine'

This use of the word *scharrel* is a case of semantic concentration, the presence of the meaning of a word that is not a formal constituent, in this case 'free range *animal*' (Meesters 2004: 52). As the last example *scharrelwijn* shows, even the notion of animal has disappeared as a necessary part of its meaning. That is, we have to assume an intermediate schema:

$$(12) \quad [[\text{scharrel}]_V [X]_{N_i}]_{N_j} \leftrightarrow [\text{ECO} [\text{SEM}_i]_j]$$

which expresses this lexicalized yet productive ecological meaning of *scharrel* when embedded in a compound.

This type of productive lexicalization can be found in many languages. A nice example comes from Maale, a North Omotic language spoken in Southern Ethiopia. The noun *nayi* 'child' has developed the general meaning 'agent', as illustrated by the following complex words (Amha, 2001: 78):

- (13) a. bayi nayi
cattle child
'one who brings cattle to the grazing area'
- b. waari nayi
goat child
'one who takes care of goats'
- c. móótsi naya
cattle.camp child

‘one who lives in a cattle camp and takes care of cattle there’

Because cattle herding is historically a task of children in the Maale-speaking society, the word for child has acquired a more general agent meaning.

In sum, we need a model of compounding in which the knowledge of individual compounds, of abstract lexical patterns, and of intermediate patterns with specific properties can be accounted for. Hence, we need a hierarchical lexicon, with constructional idioms expressing intermediate generalizations.⁴

3. *Headedness variation*

It is clear that Williams’ Right Hand Rule (Williams, 1981) cannot be a rule in the sense of a universal since many languages have left-headed compounds. Hence, one might consider the position of the head as a morphological parameter. For instance, Germanic languages may be qualified as right-headed, and Romance languages such as Italian as left-headed (Scalise, 1984, Scalise, 1992). Another example of a left-headed language is Maori (Bauer, 1993). The problem for such a parameter approach is that languages may have both left-headed and right-headed compounds. This is the case for Romance languages, and for Chinese and Japanese. Vietnamese has left-headed native compounds, and right-headed compounds borrowed from Chinese. In Javanese, compounds are left-headed except for some right-headed compounds of Sanskrit origin (Bauer, 2009: 349).

Italian and Spanish also have sets of right-headed compounds, even though the default position of compound heads is the left position. Examples of such compounds in Spanish are the following:

- (14) auto-escuela, ‘car school’, cine-club ‘cinema club’, eco-sonda ‘echo sounder’, tele-novela ‘television novel’, video-arte ‘video art’

In the framework of construction morphology we can analyse these words as compounds. We then assume constructional idioms such as

- (15) [auto_i[x]<sub>N_j]_{N_k} ↔ [SEM_j with relation R to SEM_i]_k
[tele_i [x]_{N_i]_{N_k} ↔ [SEM_i with relation R to SEM_i]_k}</sub>

instead of a general abstract template for right-headed compounds. By lexically specifying the left constituent of these compound schemas we express that the class of right-headed compounds in Spanish is restricted to compounds that begin with a word of a closed set.

Mandarin Chinese is reported to have both left-headed and right-headed compounds (Ceccagno and Basciano, 2009, Packard, 2000). Right-headed compounds have either a noun or a verb in head position. In the case of verbal compounds, the non-head functions as a modifier of the verb. Verbal compounds are left-headed, however, if the non-head functions as an argument of the verb. The following examples illustrate these patterns (Ceccagno and Basciano, 2009: 485):

- | | | | |
|------|--------------|--|---|
| (16) | right-headed | [[dú] _N [fàn] _N] _N | drug-criminal ‘drug criminal’ |
| | | [[hán] _N [shòu] _V] _V | letter-sell ‘order by mail’ |
| | left-headed | [[jìn] _V [dú] _N] _V | prohibit-poison ‘ban the sale and abuse of drugs’ |

In other words, generalizations about the position of the head must be made in terms of the corresponding semantic structure. In attributive compounds the head is on the right, whereas in compounds with a verb-argument structure the head is on the left. Such generalizations can be expressed by morphological schemas, which by definition express the correlation between the form and meaning of complex words.

There are more languages whose compounds can be either right- or left-headed. The language Biak, an Austronesian language spoken on New Guinea, has left-headed NN compounds, in which the head position is filled by nouns like *man* ‘man, bird-like entity’ or *in* ‘female person, fish-like entity’. There are also right-headed NN compounds, however. In that case, the compound expresses a specific semantic pattern, either ‘N₂ is part of N₁’, or ‘N₂ is a product of N₁’ (van den Heuvel, 2006: 91-93):

- | | | |
|------|----|--------------------|
| (17) | a. | randip-vukór |
| | | pig-head |
| | | ‘head of a pig’ |
| | b. | ai-snáw |
| | | tree-branch |
| | | ‘branch of a tree’ |

Thus, besides an abstract schema for Biak left-headed NN compounds, we need an additional one for specific semantic types of right-headed compounds. In this respect, Biak is similar to Chinese

for which language we also observed a correlation between semantics and position of the head.

Once more, we see that the idea of a hierarchical lexicon with subschemas expressing intermediate generalizations is essential for a proper account of patterns of word formation.

4.. *From compounding to derivation*

When a constituent of a compound acquires a ‘bound’ meaning, this does not necessarily lead to that meaning being an abstract, grammatical one, as we saw above. This conclusion is supported by the existence of lexical affixes in some Amerindian languages ((Gerds, 1998, Mithun, 1997, Mithun, 1999). Salishan and Wakashan languages have lexical suffixes, that is, suffixes with a specific, non-grammatical meaning (Gerds, 1998). Mithun reports that Spokane, a Salishan language has about 100 lexical suffixes that are similar to noun roots except that they do not occur as independent words. Yup’ik has more than 450 verb-like derivational suffixes. Most of the Yup’ik verbal suffixes have verbal roots as counterparts, with differences in form correlating with differences in meaning (Mithun, 1999: 48-56, Mithun, 2009a). The Amerindian language Bella Coola also has a set of lexical suffixes (Saunders and Davis, 1975).

For the Athapaskan language Slave, Rice reports that some nouns only appear as parts of compounds, not independently. Yet, such nouns can be used productively for coining new compounds. This applies, for instance, to the bound noun *teh* ‘water’ (Rice, 2009: 544). This suggests a compound schema with its head position filled lexically with the root *teh*. Similar facts obtain for Mohawk (Mithun, 2009b: 580).

Mithun points out that “a historical origin in compounding accounts well for the special properties of lexical affixes” (Mithun, 1999: 55), a position also defended by Carlson: “productive compounding, particularly of nominal objects and locatives led to the set of bound morphemes referred to as lexical affixes” (Carlson, 1990: 69). Lexical suffixes in Spokane developed from right members of compounds, and lexical prefixes from left members (Carlson, 1990: 78). The compound origin of some of these Spokane suffixes can also be concluded from the fact that the suffix begins with a linking element [ɬ] or a nominalizing element [s] that does not occur at the beginning of the corresponding independent root.

Lexical prefixes also occur in Japanese. As observed by Kageyama “Japanese has a far richer stock of prefixes than English (Kageyama, 1982: 226). For instance, there is a substantial set of prefixes with an adjectival meaning such as:

- (18) dai- ‘big’, oo- ‘big’, tyuu- ‘middle’, syoo- ‘small’, ko- ‘small’, koo- ‘high’tei- ‘low’, sin- ‘new’, ko- ‘old’, huru- ‘old’, zyuu- ‘heavy’, kei- ‘light’, tyoo- ‘long’, tan- ‘short’, kyuu- ‘sudden’, bi- beautiful’, aku- ‘bad’, koo- ‘good’

The existence of ‘bound’ compound constituents or lexical affixes receives a straightforward interpretation in a lexicon with morphological schemas that express generalizations about subsets of compounds that share one of the constituents, that is constructional idioms (schemas with some slots lexically specified) at the compound level. The bound nature of a constituent is expressed by this constituent not being co-indexed with an independent lexeme in the lexicon. Hence, the meaning of such constituents has to be specified in the corresponding constructional idiom. The origin of such lexical suffixes can thus be explained by the assumption that particular lexemes can ‘survive’ in compound schemas in which they occupy a slot, even though the corresponding lexeme got lost.

A related problem concerns the status of the morpheme *out* in English verbs such as:

- (19) outbid, outperform, outplay, outrank, outstay,

The morpheme *out* when combined with a verb has acquired (as one of its meanings) the meaning of excess, or more precisely ‘to exceed someone else in V-ing’, where V denotes the base verb. The morpheme *out-* might be considered a prefix here because of this specific meaning of *out* when combined with verbs. Indeed it is often referred to as a prefix. The same holds for a morpheme like *over* that has three meanings when combined with verbs among which the meaning of excess (Lieber, 2004: 130):

- (20) locational: overlap, overfly, overturn
completive: override, overrun
excess: overbid, overburden, overindulge

Lieber observes that such prefixes that correspond to a lexeme exhibit much more polysemy than derivational affixes, which do not have such a counterpart and tend to have one abstract meaning. Therefore, Lieber proposes the hypothesis that “prefixal *over-* is nothing more than a bound version of prepositional *over*” (Lieber, 2004: 129). This implies that such verbs are verbal compounds but that a specific subschema is necessary to express the specific meaning of *out* and *over* in combination with verbs.

In sum, what we see here, is on the one hand a form of lexicalization, words receiving specific interpretations when embedded in complex words, and on the other hand the recurrence of such lexicalized bound meanings in new words of the same type.⁵ This combination of lexicalization and productivity can therefore be interpreted as signaling the existence of constructional idioms, schemas with partially pre-specified constituents and corresponding meanings. Thus, we may assume the following constructional idiom for verbs like *outbid*:

(21) $[[\text{out}]_{\text{Adv}} [\text{x}]_{\text{vi}}]_{\text{Vj}} \leftrightarrow [\text{to exceed someone/thing in SEM}_i]_j$

In this schema, I assign the label Adv(erb) to *out*, not the label preposition because precisely those prepositions that allow for being used adverbially that can be used in compounding. For instance, preposition such as *at* and *between* are not used adverbially, and do not appear in verbal compounds.

The advantage of assuming subschemas is that we do not have to introduce a special category like semi-affix or affixoid for these phenomena. Subschemas suffice to express that speakers are able to make subgeneralizations about subsets of compounds words, and thus create new words in which the lexicalized meaning of a subconstituent of a complex word can be used productively.

Such subpatterns are a potential source of new derivational suffixes when the relation with the corresponding independent lexeme is no longer felt, due to semantic change, and when the lexeme gets lost. For instance, the English suffixes *-ful*, as in *beautiful*, and *-able*, as in *washable*, are no longer felt to be related to the lexemes *full* and *able*. A suffix like *-hood* derives historically from a lexeme with the meaning ‘quality’. The Dutch suffix *-lijk* and its English counterpart *-ly* derive from the noun *leik* ‘body’, and the suffix *-dom* (a suffix in both Dutch and English) derives from a lexeme for ‘dominion’. The best known case for Romance languages is the adverbializing suffix *-mente* (French form *-ment*) which derives from the Latin noun *mens* ‘mind’ in its ablative form, as in *clara-mente* ‘with a clear mind, in a clear way’.

This rise of derivational morphemes is often qualified as grammaticalization (Aikhenvald, 2007: 58), since these morphemes have become affixes. Yet, if situated at the endpoint of grammaticalization, we expect these morphemes to have abstract grammatical properties, whereas a morpheme like *-dom* still has a rather specific meaning. Hence, it seems that there is a cline for such bound morphemes ranging from a more lexical to a more grammatical meaning.

Thus, the rise of bound meanings for lexemes embedded in complex words, and the change of lexemes into affixes shows the necessity of assuming morphological subschemas that account for the bound interpretations of lexemes, and for the possibility of such changes.

5. Allomorphy

Lexemes may exhibit systematic stem allomorphy when embedded in compounds. Such systematic stem allomorphy can be captured by means of schemas that express the relevant generalizations. In this section I discuss a number of these regularities for Dutch compounds.

A first example is the allomorphy of the lexeme MEDE ‘with, together’. This word is used as preposition, postposition, as particle, and as part of a compound, with different shapes:

- (22) preposition: met [mɛt]
 postposition and particle: mee [me:]
 first part of a compound: mede [me:də]

Historically the three forms are related as follows: the long and original form is *mede*. The short form *mee* is the effect of a historical phonological process of *de*-deletion. The form *met* is the effect of schwa apocope followed by word-final devoicing of obstruents, and vowel shortening,

The long form *mede* occurs in a few (archaic) particle verbs: *mede-dingen* (but also *mee-dingen*) ‘to compete’, and *mededelen / meedelen* ‘to inform’. Otherwise, the particle form is identical to that of the postposition, *mee*. That is, *mede* is only productively used within compounds, and has the specific meaning ‘shared with, co-, fellow-’:

- (23) a. [[mede]_{ADV} N_i]_{N_j} ↔ [SEM_i shared with others]_j
 b. mede-beslissing ‘co-decision
 mede-bewoner ‘fellow occupant’
 mede-broeder ‘fellow brother’

The use of *mee* as a particle is illustrated by the following particle verbs; it is a productive category:

- (24) mee-bidden ‘to join in praying’
 mee-denken ‘to join in thinking’

mee-drinken ‘to join in drinking’
mee-eten ‘to join in eating’

We might qualify *mede* as a prefix, but there is a clear lexical relation with the adverbial particle *mee*, both in form and meaning. A schema as given in (23a) is therefore to be preferred to a prefixal interpretation: *mede* is a form of a lexeme, but a specific allomorph selected by nominal compounds, with a correlating compound-determined meaning.

The existence of schema (23a) will block the formation of [Adv N] compounds of the form *mee* + N. Such compounds are indeed avoided in Dutch, even though the compound schema [Adv N]_N is productive in Dutch; this blocking effect is illustrated by the examples in (25):

- (25) mede-bewoner / *mee-bewoner ‘co-occupant’
mede-gelovige / *mee-gelovige ‘fellow believer’
mede-klinker / *mee-klinker ‘consonant’

A noun like *mee-eter* ‘one who joins in eating’ is possible, however, since it is derived from the established particle verb *mee-eten* ‘to join in eating’, but it has a meaning different from that of the compound *mede-eter* ‘fellow diner’. Therefore, the two forms are not synonymous, and hence do not compete, and blocking is not involved.

The required blocking of the incorrect forms in (25) follows from Panini’s principle: the schema for Adv N compounds with *mede* specified as the Adverb position is more specific than the general schema [Adv N]_N. Hence, it will override the general schema, and block its application to the adverb *mee*.

In some cases we find semantic differences between the allomorphs of a noun. The word *eer* [e:r] ‘honour’ has a long allomorph *ere* [e:rə] that has an archaic flavour when used as an independent word. As the first constituent of a compound, we find both *eer-* and *ere-*. However, if the meaning to be expressed is ‘honorary’, one must always use the long form *ere-*:

- (26) a. eer-wraak ‘honour-revenge, revenge for the protection of family honour’
eer-betoon ‘honour-show, tribute’
b. ere-lid ‘honour-member, honorary member’ / *eer-lid
ere-voorzitter ‘honour-chairman, honorary chairman’ / *eer-voorzitter
ere-doctor ‘honour-doctor, honorary doctor’ / *eer-doctor

This implies that the following schema has to be assumed:

$$(27) \quad [[ere]_N[x]_{N_i}]_{N_j} \leftrightarrow [\text{honorary SEM}_i]_j$$

This schema will block the insertion of the short noun allomorph *eer* ‘honour’ in a general $[NN]_N$ schema with the specific ‘honorary’ meaning of the lexeme *eer* since schema (27) is more specific than the general schema for NN compounds.

Allomorphy patterns thus provide additional evidence for the necessity of subschemas for the coinage of complex words.

6. *Analogy or schema?*

A old debate in the analysis of newly coined complex words is whether this has to be considered as a matter of analogy, or as the result of using symbolic schemas that generalize across sets of existing complex words. The implication of assuming a hierarchical lexicon with different levels of abstraction is that this is not a matter of ‘either/or’; there is analogical word formation, based on an individual model word, but there is also word formation based on schemas. These schemas may, however, differ in their degrees of abstractness. Hence, specific sets of existing complex words may play a role. Moreover, it is not the case that all language users make the same subgeneralizations. Schemas are based on lexical knowledge, and this type of knowledge varies from speaker to speaker (Langacker, 1991, Taylor, 2002, Tuggy, 2007).

The use of schemas for word formation patterns implies a symbolic approach to representing linguistic knowledge. I am fully aware of the fact that there are models of morphology that try to do away with symbolic representation of morphological knowledge, models in which analogy to existing words and memory-based learning play a central role (Daelemans, 2002, Keuleers and Daelemans, 2007, Keuleers et al., 2007). In such models the notion ‘analogy’ has received an elaborate sophistication. They have been developed for inflectional processes in which a choice has to be made between different inflectional endings, as in the case of Dutch plural nouns. The correct inflectional endings are computed by measuring the degree of similarity between the input word and the set of words in the lexicon, and selecting the inflectional form that corresponds to that of the most similar word(s) found.

Similar selection problems must be faced when selecting the proper linking element in a newly coined Dutch compound. The linking element to be selected is that of those existing compounds to which the new compound is most similar. In a study of Dutch compounding, Krott

showed that analogical modeling is able to cover the selection of linking elements in an adequate way (Krott, 2001).. Similarly, Plag argued that analogy plays a role in computing the correct stress pattern of English compounds (Plag, 2006).

In my opinion, these findings are not in conflict with the kind of word formation schemas that I proposed above. The creation of a new compound involves two stages. In the first stage of language production, it has to be decided how a particular semantic content is going to be expressed (Levelt, 1989). One of the options is to select a particular word for that content. The range of possible morphological forms for the expression of semantic content is specified by a set of hierarchically ordered schemas in the lexicon. One of these schemas might be selected, and open positions filled in with lexemes. For instance, if we want to express the concept ‘main’ in Dutch, speakers of Dutch know that they can use the schema $[[hoofd]N]_N$ with the meaning ‘main N’ for this purpose. This kind of knowledge must therefore be assumed to exist in addition to a set of existing compounds of that form.

In the second stage of language production, the exact form of the compound constituents must be computed. If a lexeme has more than one form – for instance since it may combine with different linking elements – the selection of a particular form may well be modeled properly by the analogical models referred to above since it is clear that the language user has access to existing compounds that share properties such as the initial constituent lexeme with the new compound. The language user who wants to speak about the food that a sheep is provided with may first decide to use an NN compound that begins with the lexeme SCHAAP ‘sheep’ as its modifier noun, and the lexeme VOER as the head lexeme. At this stage, use is made of the abstract schema for NN compounds that specifies that the right N is the head, or a relevant subschema thereof. Then the question arises whether the correct form of the lexeme SCHAAP is *schaap*, *schapen-*, or *schaaps-*. It is at that stage that analogical modeling can make predictions as to which allomorph is preferably chosen (the form *schapenvoer* should be predicted as being the most probable one since *schapen* is the most frequently used allomorph, Similarly, the choice of the correct stress pattern for a new English compound at stage 2 may be modeled analogically, after the selection of a particular compounding scheme for coining a new compound has taken place.

As pointed out by Baayen, “the symbolic approach in which paradigmatic structure provides a similarity space over which probabilities are defined provides an excellent level of granularity for understanding the role of probability in language production” (Baayen, 2003: 63). A hierarchical lexicon with different levels of abstractness and generalization, as outlined in this paper, expresses this paradigmatic structure, and thus the relevant similarity space.

In sum, a hierarchical lexicon with constructional idioms for subsets of compounds is an essential tool in modeling the regularities in the semantic interpretation and formal make-up of compounds.

Notes

1. For discussion of the notion ‘analogy’ with respect to word formation, see (Becker, 1990, Becker, 1994, Hüning, 1999).
2. The existence of constituent families is confirmed by psycholinguistic evidence, in particular through the family size effect: the larger the size of a constituent family of a word, the faster it will be retrieved (Baayen, 2003, Schreuder and Baayen, 1997).
3. Phrases such as *dol- en dolblij* may be derived from *dolblij and dolblij* through the rule of prosodic gapping that deletes one of two identical prosodic words (Booij, 1985). However, prosodic gapping is normally optional, whereas omission of the first of two identical constituent is obligatory in these cases of emphatic repetitive coordination. This implies that this instantiation of gapping is to be considered as a subconstruction of gapping; the obligatoriness of the gapping and the formal identity of the coordinated adjectives corresponds with an emphatic meaning.
4. Lieber suggests a constructional idiom for a subset of compounds for semantic reasons as well: compounds of the form [*media Xion*] tend to be subject-oriented, with *media* playing the role of subject, as in *media competition*, whereas in most synthetic compounds the first constituent functions as an object (Lieber, 2009: fn 7).
5. This point of view is also defended in (Brinton and Traugott, 2005: 129).

REFERENCES

- Aikhenvald, Alexandra J. 2007. Typological distinctions in word-formation. In *Language typology and syntactic description, Vol III: Grammatical categories and the lexicon*, ed. Timothy Shopen, 1-65. Cambridge: Cambridge University Press.
- Amha, Azeb. 2001. *The Maale language*. Leiden: University of Leiden, Research School of Asian, African, and Amerindian Studies.
- Ascoop, Kirstin. 2005. Affixoidhungrig? Skitbra! Status und Gebrauch von Affixoiden im Deutschen und Schwedischen. *Germanistische Mitteilungen* 65:17-28.
- Ascoop, Kristin, and Leuschner, Torsten. 2006. Affixoidhungrig? Skitbra! Comparing affixoids in Swedish and German. *Sprachtypologie und Universalienforschung* 59:241-252.

- Baayen, R. Harald. 2003. Probabilistic approaches to morphology. In *Probabilistic linguistics*, eds. Rens Bod, Jennifer Hay and Stefanie Jannedy, 229-287. Cambridge Mass. / London Engl.: MIT Press.
- Bauer, Laurie. 2009. Typology of compounds. In *The Oxford handbook of compounding*, eds. Rochelle Lieber and Pavol Stekauer, 343-356. Oxford: Oxford University Press.
- Bauer, Winifred. 1993. *Maori*. London / New York: Routledge.
- Becker, Thomas. 1990. *Analogie und morphologische Theorie*. München: Wilhelm Fink Verlag.
- Becker, Thomas. 1994. Back-formation, cross-formation, and "bracketing paradoxes". In *Yearbook of Morphology 1993*, eds. Geert Booij and Jaap van Marle, 1-26. Dordrecht: Kluwer.
- Booij, Geert. 1985. Coordination reduction in complex words: a case for prosodic phonology. In *Advances in non-linear phonology*, eds. Harry van der Hulst and Norval Smith, 143-160. Dordrecht: Foris Publications.
- Booij, Geert. 2005. Compounding and derivation: evidence for construction morphology. In *Morphology and its demarcations*, eds. Wolfgang U. Dressler, Dieter Kastovsky, Oskar E. Pfeiffer and Franz Rainer, 109-132. Amsterdam / Philadelphia: John Benjamins.
- Booij, Geert. 2007. Construction morphology and the lexicon. In *Selected proceedings of the 5th Décembrettes: Morphology in Toulouse*, eds. Fabio Montermini, Gilles Boyé and Nabil Hathout, 34-44. Somerville: Cascadilla Press.
- Brinton, Laurel J., and Traugott, Elizabeth Closs. 2005. *Lexicalization and language change*. Cambridge UK: Cambridge University Press.
- Carlson, Barry. 1990. Compounding and lexical affixation in Spokane. *Anthropological Linguistics* 32:69-82.
- Ceccagno, Antonella, and Basciano, Bianca. 2009. Sino-Tibetan: Mandarin Chinese. In *The Oxford handbook of compounding*, eds. Rochelle Lieber and Pavol Stekauer, 478-490. Oxford: Oxford University Press.
- Daelemans, Walter. 2002. A comparison of analogical modeling of language to memory-based language processing. In *Analogical modeling*, eds. Royal Skousen, Deryle Lonsdale and Dilworth B. Parkinson, 157-179. Amsterdam / Philadelphia: Benjamins.
- Dalton-Puffer, Christiane, and Plag, Ingo. 2000. Category-wise, some compound-type morphemes seem to be rather suffix-like: on the status of -ful, -type, and -wise in present-day English. *Folia Linguistica* 34:225-244.
- Gerdts, Donna B. 1998. Incorporation. In *The handbook of morphology*, eds. Andrew Spencer and Arnold Zwicky, 84-104. Oxford U.K.: Blackwell.
- Hüning, Matthias. 1999. *Woordensmederij. De geschiedenis van het suffix -erij*. Utrecht: LOT.
- Hüning, Matthias. 2000. Monica en andere gates. Het ontstaan van een morfologisch procédé. *Nederlandse Taalkunde* 5:121-132.
- Jackendoff, Ray. 2002. *Foundations of language*. Oxford: Oxford University Press.
- Kageyama, Taro. 1982. Word formation in Japanese. *Lingua* 57:215-258.

- Keuleers, Emmanuel, and Daelemans, Walter. 2007. Memory-based learning models of inflectional morphology: a methodological case study. *Lingue e Linguaggio* 6:151-174.
- Keuleers, Emmanuel, Sandra, Dominiek, Daelemans, Walter, Gillis, Steven, Durieux, G., and Martens, E. 2007. Dutch plural inflection: The exception that proves the analogy. *Cognitive Psychology* 54:283-318.
- Krott, Andrea. 2001. *Analogy in morphology: The selection of linking elements in Dutch compounds*. Nijmegen: Max Planck Institut für Psycholinguistik.
- Langacker, Ronald. 1991. *Foundations of cognitive grammar. vol. 2. Descriptive applications*. Stanford Calif.: Stanford University Press.
- Leuschner, Torsten, and Decroos, Nancy. 2008. Wortbildung zwischen System und Norm. Affixoiden im Deutschen und im Niederländischen. *Sprachwissenschaft* 33:1-34.
- Levelt, Willem J. M. 1989. *Speaking: from intention to articulation*. Cambridge Mass.: MIT Press.
- Lieber, Rochelle. 2004. *Morphology and lexical semantics*. Cambridge: Cambridge University Press.
- Lieber, Rochelle. 2009. On the lexical semantics of compounds. In *Compounding*, eds. Irene Vogel and Sergio Scalise. Amsterdam / Philadelphia: Benjamins.
- Malkiel, Yakov. 1978. Derivational categories. In *Universals of human language. Volume 3. Word structure*, ed. Joseph H. Greenberg, 125-149. Stanford Calif.: Stanford University Press.
- Mithun, Marianne. 1997. Lexical affixes and morphological typology. In *Essays on language function and language type*, eds. Joan Bybee, John Haiman and Sandra Thompson, 357-372. Amsterdam / Philadelphia: Benjamins.
- Mithun, Marianne. 1999. *The languages of Native North America*. Cambridge: Cambridge University Press.
- Mithun, Marianne. 2009a. Compounding and incorporation. In *Compounding*, eds. Sergio Scalise and Irene Vogel. Amsterdam / Philadelphia: Benjamins.
- Mithun, Marianne. 2009b. Mohawk. In *The Oxford handbook of compounding*, eds. Rochelle Lieber and Pavol Stekauer, 564-588. Oxford: Oxford University Press.
- Packard, Jerome. 2000. *The morphology of Chinese. A linguistic and cognitive approach*. Cambridge UK: Cambridge University Press.
- Plag, Ingo. 2006. The variability of compound stress in English: structural, semantic, and analogical factors. *English Language and Linguistics* 10:143-172.
- Ramat, Paolo. 2001. Degrammaticalization or transcategorization? In *Naturally! Linguistic studies in honour of Wolfgang Ulrich Dressler presented on the occasion of his 60th birthday*, eds. Chris Schaner-Wolles, John Rennison and Friedrich Neubarth, 393-401. Torino: Rosenberg & Sellier.
- Rice, Keren. 2009. Athapaskan: Slave. In *The Oxford handbook of compounding*, eds. Rochelle Lieber and Pavol Stekauer, 542-563. Oxford: Oxford University Press.
- Saunders, Ross, and Davis, Philip W. 1975. Bella Coola lexical suffixes. *Anthropological Linguistics* 17:154-189.
- Scalise, Sergio. 1984. *Generative morphology*. Dordrecht: Foris.
- Scalise, Sergio. 1992. Compounding in Italian. *Rivista di Linguistica* 4:175-199.
- Schreuder, Rob, and Baayen, R. Harald. 1997. How complex simplex words can be. *Journal of Memory and Language* 37:118-139.

- Taylor, John R. 2002. *Cognitive grammar*. Oxford: Oxford University Press.
- Tuggy, David. 2007. Schematicity. In *The Oxford handbook of cognitive linguistics*, eds. Dirk Geeraerts and Hubert Cuyckens, 82-116. Oxford: Oxford University Press.
- van den Heuvel, Wilco. 2006. *Biak. Description of an Austronesian language of Papua*. Utrecht: LOT.
- Williams, Edwin. 1981. On the notions 'lexically related' and 'head of a word'. *Linguistic Inquiry* 12:245-274.

Terms for index

allomorphy

analogy

constructional idiom

construction morphology

derivation

grammaticalization

headedness

lexical affix

lexicalization

morphological schema

polysemy

semantic concentration

subschemas