The role of second order schemas in the construction of complex words

Geert Booij
University of Leiden
G.E.Booij@hum.leidenuniv.nl

Francesca Masini
University of Bologna
francesca.masini@unibo.it

Abstract

Morphology is the study of the systematic relationship between the form and meaning of complex words. Therefore, it is a central task of morphology to provide a proper account of how the meanings of complex words are computed. One straightforward approach would be to assume that the computation of complex words is ruled by Fregean compositionality. The latter, however, has been claimed to be too narrow, since both syntactic and morphological constructions may exhibit specific holistic semantic properties that cannot be derived from their constituents or from general patterns of combination (Booij 2010; Goldberg 1995; Goldberg 2006; Jackendoff 2013). In the article we address a related problem, i.e. the fact that the meaning of a complex word may derive from that of another linguistic construct (be it a word or a phrase) that is not a building block of that complex word. We illustrate this point by providing data from different languages and we claim that this type of violation of Fregean compositionality can be accounted for by means of "second order schemas", i.e. sets of two or more paradigmatically related constructional schemas.

1. Introduction: compositionality problems

Morphology is the study of the systematic relationship between the form and the meaning of complex words. Therefore, it is a central task of morphology to provide a proper account of how the meanings of complex words are computed. A first approach is to use the principle of compositionality, usually attributed to Gottlob Frege (1892), as a starting point: "The meaning of a sentence is a function of the meanings of the words in it and the way they are combined syntactically" (Dowty 2007: 23). We might assume a parallel definition for the computation of the meaning of complex words: "The meaning of a (complex) word is a function of the meanings of its constituents and the way they are combined morphologically". However, it has become clear that, both for syntactic and morphological constructs, these Fregean definitions of compositionality are too narrow.

First, syntactic and morphological constructions may exhibit specific holistic semantic properties that cannot be derived from their constituents and the way they are combined (Booij 2010; Goldberg 1995, 2006; Jackendoff 2013). For instance, in present-day English we find reduplicative compounds of the type *salad-salad* 'real salad' in

which the occurrence of two identical constituents expresses a specific constructional meaning: a salad-salad is a "(proto)typical" salad (Ghomeshi et al. 2004). The same holds for Dutch and Italian. For instance, the Dutch reduplicative compound *vakantie-vakantie* 'vacation-vacation' denotes a real vacation (where people actually relax, with no work still to be done). Similarly, in Italian an expression such as *attore-attore* 'actor-actor' may refer to a person who is a true, full-fledged and possibly famous actor. The general arrangement of these constituents is that of compounding. However, this compounding arrangement does not predict the more specific meaning of these reduplicative compounds. This meaning component of prototypicality or intensity is evoked by the copying configuration of these compounds, and hence it is a holistic property of this type of compounding. This observation is not necessarily to be interpreted as a violation of Fregean compositionality, but it does imply that the general principle of compositionality does not suffice to compute the meaning of such complex words.

Second, and this is the specific problem that we want to address in this article, the meaning of a complex word may be compositionally derived by referring to a linguistic construct (word or phrase) that does not form a subconstituent of that complex word. Hence, this is a strict violation of Fregean compositionality, which is defined in purely syntagmatic terms. Let us illustrate this problem by means of a phenomenon in Modern Greek, which is normally referred to as the Bare Stem Constraint (Ralli 2013: 133ff., 247). In Modern Greek complex adjectives can be 'derived' from corresponding AN sequences which are phrasal in nature, but have the status of lexical items:

(1)   The Bare Stem Constraint
        Modern Greek AN phrase   >   Complex (relational) adjective
        psixrós pólemos   *psixr-o-polem-ik-ós*
        'cold war'   'cold-war like'

What we see here is that the conventionalized meaning of the phrase *psixrós pólemos* 'cold war' recurs in the meaning of the corresponding adjective. Yet, the proper form of the adjective is not *psixrós-polem-ik-ós* (with the stem form of the phrase *psixrós pólemos* as its base), but a compound with the structure *psixr-o-polem-ik-ós*, with two stems, and the usual linking element *-o-* of Greek compounds. The Bare Stem Constraint says that in Modern Greek words cannot appear in complex words in their inflected form, the inflectional ending has to be omitted. Hence, for the proper semantic interpretation of these complex adjectives we have to refer to the meaning of corresponding lexical phrases. This correspondence is paradigmatic in nature, because the phrases are indeed not building blocks (structurally speaking) of the corresponding adjectives, although their lexical stems do occur in the corresponding compounds. As we will see in Section 4, something very similar happens in Russian.

The claim that we will defend in this article is that this type of violation of Fregean compositionality can be accounted for by means of "second order schemas". A second order schema is a set of two or more paradigmatically related schemas. This term has been introduced by Nesset (2008) in his analysis of allomorphy patterns in Russian inflection. It has also been argued for in Booij (2010) and Kapatsinski (2013) for the domain of inflection.

In Section 2 we will give a general motivation of the use of second order schemas for a proper account of word formation patterns, including their semantics, in the framework of Construction Morphology (CxM) (Booij 2010), mainly on the basis of data from Dutch. The use of second order schemas presupposes two basic claims of CxM: (i) the necessity of constructional schemas, and (ii) the role of paradigmatic relationships in accounting for the structure of lexical knowledge and word formation patterns. In the remaining sections we will argue that second order schemas can be used to explain the semantics of complex words that are related paradigmatically to phrasal lexemes. Section 3 discusses cases of bracketing paradoxes in Italian. Section 4 deals with Russian "squeezed" phrasal lexemes of various sorts, and Section 5 with nominalizations of Dutch particle verbs. Section 6 draws some conclusions about the form-meaning relationships in complex words.

2. The need for second order schemas

The necessity of second order schemas for the analysis of certain word formation patterns is argued for in Booij (2010: 31-36). An important argument for the assumption of second order schemas in the domain of word formation is provided by patterns of affix replacement (Booij 2002, 2010). For instance, in Dutch, deverbal nouns may be derived from verbal stems in *-eer* by replacing this suffix with the suffix *-atie*. Alternatively, deverbal nouns may be formed by means of the suffix *-ing*. Hence, we have the following pattern:

(2)   *verb*   *deverbal noun*
        constat-eer 'observe'   constat-er-ing / constat-atie 'observation'
        reden-eer 'reason'   reden-er-ing / reden-atie 'reasoning'
        situ-eer 'situate'   situ-er-ing / situ-atie 'situation'

Similar facts concerning English are discussed in Aronoff (1976: 88-98). An example is the relation between the verb *nomin-ate* and its nominal derivative *nomin-ee*. Aronoff's solution for the affix replacement pattern was to assume truncation rules that delete a suffix before another suffix. Thus, a concatenative morphology interpretation of this pattern could be maintained. However, this account has a serious drawback: truncation rules are an ad hoc mechanism (the concatenation of suffixes is possible in other cases), only introduced to incorporate affix replacement in a model of word formation that only allows for concatenative morphology.

Another example of affix replacement can be found in English (Booij 2010: 31-36). Consider the following English word pairs in *-ism* and *-ist*:

(3)   altru-ism   altru-ist
        aut-ism   aut-ist
        bapt-ism   bapt-ist
        commun-ism   commun-ist
        pacif-ism   pacif-ist

Even though they have no corresponding base word, the meaning of one member of a pair can be defined in terms of that of the other member. In particular, the meaning of the word in -*ist* can often be paraphrased as 'person with the ability, disposition, or ideology denoted by the word in -*ism*'. Hence, the following paradigmatic relationship can be defined for these two schemas:

(4) $< [x\text{-}ism]_{Ni} \leftrightarrow SEM_i > \approx < [x\text{-}ist]_{Nj} \leftrightarrow [\text{person with property Y related to } SEM_i]_j >$

where SEM_i represents the set of meanings {ABILITY, DISPOSITON, IDEOLOGY}. The symbol $\approx$ is used in Booij (2010) to indicate a paradigmatic relationship between two constructional schemas. Thus, an altruist has a disposition for altruism, and a pacifist adheres to the ideology of pacifism. The paradigmatic relationship between these two schemas may lead to the coining of new words. For instance, if we know what *determinism* is, we can easily coin the word *determinist*, and then we know that this word denotes a person adhering to determinism. The same holds for nouns in -*ist* with a lexeme as their base, such as *Marxist* and *socialist*. A Marxist is an adherent of Marxism and not necessarily a follower of Marx, since Marxism as a doctrine encompasses more than the ideas of Marx (in fact, Marx himself declared that he was not a Marxist). Similarly, a socialist is not necessarily a social person, but an adherent of the ideology of socialism.

Schema (4) does not imply that all nouns in -*ist* correspond to a noun in -*ism*, only those with the meaning specified in (4) do. For instance, for nouns in -*ist* with the meaning 'practitioner of', such as *linguist* and *semanticist*, there are no corresponding nouns *lingu-ism* or *semantic-ism*. In the domain of sciences and their practitioners, there are other second order schemas involved, with correlations such as (roughly) $<Xics \leftrightarrow X>$ (*linguistics-linguist*) or $<Xic\text{-}s \leftrightarrow Xc\text{-}ist>$ (*semantics-semanticist*). That is, there are several paradigmatic patterns involved in the interpretation of nouns in -*ist*. In principle, this does not exclude the possibility (to be checked, though) to posit a more abstract (in terms of both form and meaning) second order schema for all -*ist* nouns, from which the various subtypes are instantiated.

In sum, we need a second order schema like (4) for an adequate account of the semantics of certain sets of words in -*ist*. The meaning of these nouns in -*ist* is not simply a compositional function of their constituent parts, but contains the meaning of a related word with the same degree of complexity.

Crucially, even though semantically the word in -*ism* is the starting point for the word in -*ist*, this does not mean that the actual order of derivation necessarily reflects this semantic asymmetry. For instance, the word *abolitionist* may have been coined before *abolitionism*. So, another advantage of paradigmatic relationships like that in (4) is that they allow for word formation in both directions.

An additional formal argument for this type of analysis is that non-native roots have all sorts of idiosyncratic allomorphy. For instance, correlated to the noun *Plato* we find *platon-ist*, *platon-ism* and *platon-ic*. That is, in complex words the allomorph of the stem is systematically *platon-*. This is accounted for by the analysis proposed here. Similarly, the allomorph *mis-* of *mit-* (as in *submit*) appears both in *submiss-ion* and *submiss-ive*, which is predicted by a second order schema for nouns in -*ion* and adjectives in -*ive*.

4

We find such systematic paradigmatic relationships across Germanic (Becker 1990, 1994) and Romance languages (Vallès 2003). In a rule-based framework such relationships would require an operation of affix replacement instead of affix concatenation. In a schema-based analysis, it suffices to state the formal and semantic correlation between two classes of words with the same degree of morphological complexity. Thus, the assumption of second order schemas provides the means for expressing paradigmatically governed generalizations regarding the semantic interpretation of complex words.

Second order schemas are also essential in order to account for what we might call overcharacterization: the selection of the proper form of bases in word formation may be dependent on the form of a paradigmatically related word that does not function as a semantic base. A well-known example is the formation of certain classes of toponyms in Dutch (Booij 1997a, 1997b, 2010). For instance, in quite a number of cases, when we coin the adjective for a certain country in Dutch, we do not add the dedicated suffix (mainly -s, sometimes -isch /is) to the name of that country, but to that of the corresponding name for the inhabitant:[1]

(5)   *country*   *inhabitant*   *adjective for country*
      Zweden 'Sweden'   Zweed 'Swede'   Zweed-s 'Swedish'
      Noorwegen 'norway'   Noor 'Norwegian'   Noor-s 'Norwegian'
      Finland 'Finland'   Fin 'Finn'   Fin-s 'Finnish'
      Denemarken 'Denmark'   Deen 'Dane'   Deen-s 'Danish'
      Italië 'Italy'   Italiaan 'Italian'   Italiaan-s 'Italian'
      Rusland 'Russia'   Rus 'Russian   Russ-isch 'Russian'
      België 'Belgium'   Belg 'Belgian'   Belg-isch 'Belgian'

This type of paradigmatically determined choice of the form of a base, with concomitant overcharacterization, also plays a role in the coinage of female inhabitant names in Dutch: the suffix -e is not added to the neutral / male inhabitant name, but to the corresponding adjective:

(6)   *inhabitant*   *adjective*   *female inhabitant*
      Zweed 'Swede'   Zweed-s 'Swedish'   Zweeds-e
      Noor 'Norwegian'   Noor-s 'Norwegian'   Noors-e
      Fin 'Finn'   Fin-s 'Finnish'   Fins-e
      Deen 'Dane'   Deen-s 'Danish'   Deens-e
      Italiaan 'Italian'   Italiaan-s 'Italian'   Italiaans-e
      Rus 'Russian   Russ-isch 'Russian'   Russisch-e
      Belg 'Belgian'   Belg-isch 'Belgian'   Belgisch-e

Similar data have been observed for French by Namer (2013). For instance, the verb *patronaliser* 'to assign (something) to the employer' has *patronal* 'employer-like' as its formal base, not *patron* 'employer'. The verb *amicaliser* can be interpreted as derived

---

[1] Not all adjectives conform to this pattern, as shown by the word series *Frankrijk* 'France' – *Fransman* 'Frenchman' – *Frans* 'French', and *Nederland* 'Netherlands' – *Nederlander* 'Dutchman' – *Nederlands* 'Dutch'.

from the adjective *amical* 'friendly', thus having the meaning 'to make friendly', but it can also mean 'to make a friend', thus deriving semantically from *ami* 'friend', as in *Amicalise-moi sur Facebook* 'Friend me on Facebook'. As Namer (2013) put it: "An additional stem to a noun may be supplied by the relational adjective the noun is related to, and whose frequency makes it easily accessible in the speaker's mental lexicon".

Such patterns of relationship with a mismatch between form and meaning can be accounted for by means of second order schemas. For instance, the relationship between toponymical adjectives and the corresponding country names in which the inhabitant names play the intermediary role of providing the form of the base, as shown in (5), is expressed by the following second order schema (Booij 2010: 35):

(7) $< [x]_{Ni} \leftrightarrow [\text{inhabitant of } j]_i > \approx < [[x]_{Ni}\text{-}(i)s]_{Ak} \leftrightarrow [\text{relating to } j]_k >$

The necessity of second order schemas is also shown by an intriguing type of word formation in Dutch, that of elative compounds in which the first part is a noun in its diminutive form. Elative compounds are adjectival compounds in which the first constituent has acquired a more abstract meaning and expresses intensity or disapproval (Hoeksema 2012). The first constituent may be a noun, verb or adjective. Here are some examples:

(8)   NA   bloed-geil   'blood-horny, very horny'
             stront-vervelend   'shit-boring, very boring'
        VA   stik-heet   'suffocate-hot, very hot'
             knetter-gek   'crack-mad, very mad'
        AA   wild-vreemd   'wild-strange, very strange'

These elative compounds have to be accounted for by a specific subschema for Dutch compounds in which the intensifying meaning is specified:

(9) $<[X \text{ } A_i]_{Aj} \leftrightarrow [\text{high degree of SEM}_i]_j>$

Words that instantiate this schema have specific formal properties as well, in particular they allow for the repetition of the first constituent, as in *bloed- en bloedgeil* 'very, very horny' (cf. Booij 2010: 55-60 for details). This subschema for adjectival compounds will dominate a number of subschemas in which the first constituent is lexically specified, as there is a specific closed class of words that can function as intensifier in these elative compounds.

The remarkable property of these compounds is that the first constituent can also appear in a diminutive form, which however expresses an even higher degree of the property denoted by the adjectival head of such compounds (Morris 2013):

(10)   bloed-je-geil   'blood-DIM-horny, very horny'
        stront-je-vervelend   'shit-DIM-boring, very boring'
        stik-je-heet   'suffocate-DIM-hot, very hot'
        knetter-tje-gek   'crack-DIM-mad, very mad'
        wild-je-vreemd   'wild-DIM-strange, very strange'

The diminutive suffix DIM (which has a number of allomorphs, among which *-je* and *-tje*) used in this way can also be inserted after prefixes such as *ultra-, über-* and *super-*, and after a cranberry morpheme[2] like *tjok* 'chock' (data from Morris 2013):

(11)   ultra-tje-kort   'very short'
       über-tje-cool   'very, very cool'
       super-tje-good   'very, very good'
       tjok-je-vol   'very full'

The complex adjectives in (11) are formally compounds consisting of a diminutive noun followed by and adjective. The Dutch diminutive suffix is a category-determining suffix and always create nouns, whatever the syntactic category of its base. The kind of diminutive nouns that we find in these elative compounds is special in that other categories than nouns are used for these diminutive forms, whereas normally this suffix only rarely takes other bases than nouns. The interpretation of these compounds as expressing a high degree of intensification crucially depends on the existence of a corresponding elative complex word. Note also that the diminutive suffix can normally not be attached to prefixes or cranberry morphemes. Therefore, we can only account for the occurrence of these diminutive elative compounds by assuming a second order schema of the following form:

(12) $<[X A_i]_{Aj} \leftrightarrow [\text{high degree of SEM}_i]_j> \approx$
       $<[X+(t)je A_i]_{Ak} \leftrightarrow [\text{high degree of SEM}_j]_k>$

Schema (12) expresses that these "diminutive compounds" express a further intensification of the high degree of the property denoted by the adjective.

In conclusion, in this section we have shown that second order schemas are necessary for a proper account of a number of word formation processes, in particular in Dutch. In the next two sections we will support this claim by considering data from Italian and Russian.

3. Bracketing paradoxes in Italian

Bracketing paradoxes are well-known cases of mismatch between form and meaning (Williams 1981, DiSciullo & Williams 1987). Let us consider those paradoxes that span morphology and syntax, exemplified by the notorious example in (13) (Spencer 1988, 1991). Here two different bracketings seem to be required, one for form (13)a and one for meaning (13)b (see Booij 2010: 139):

(13)   transformational grammar > transformational grammarian
       a.   form: $[[\text{transformational}]_A [\text{grammar-ian}]_N]_{NP}$
       b.   meaning: $[[\text{transformational grammar}]_{NP} \text{-ian}]_N$

---

[2] Cranberry morphemes are morphemes that only appear as stems in a complex word, as *mer-* in *mermaid*.

The problematic nature of this example lies in the fact that, formally, the suffix *-ian* attaches to *grammar*, but semantically it has scope on the whole phrase *transformational grammar*, since a *transformational grammarian* is a grammarian that works within the framework of transformational grammar and not a grammarian who is transformational. Hence, either we assume that *-ian* takes a whole phrase as input, against the Lexical Integrity Hypothesis and the No-Phrase Constraint (Botha 1984), or we end up with a problematic mismatch. Spencer (1988) proposed to solve the dilemma by invoking the paradigmatic dimension of this kind of formation, based the "proportional analogy", which presupposes the existence of lexical phrases (e.g. *transformational grammar*) listed in the lexicon.

In Italian, as noticed by Bisetto & Moschin (2010), things are even more complicated, since in cases comparable to (13) the affix attaches to the left (head) constituent, thus splitting the original NA phrase (example from Bisetto & Scalise 1991: 36, ft. 3):

(14)   flauto barocco   →   flaut-*ista* barocco
       flute baroque   flute-ist baroque
       'baroque flute'   'baroque flutist'

A *flautista barocco* is not a flutist that is baroque (although it could be), but a flutist who plays the baroque flute.

This is far from being an ad hoc formation process. In fact, it is quite productive in a variety of domains (see Virgillito 2010):

(15)   a.   chitarra acustica   →   chitarr*ista* acustico
            'acoustic guitar'   'acoustic guitarist'
       b.   violino classico   →   violin*ista* classico
            'classical violin'   'classical violinist'
       c.   economia politica   →   econom*ista* politico
            'political economy'   'political economist'

A similar mechanism is at work in the pairs in (16), where we do not have suffixation (on the head noun), but a case of affix replacement (see Section 1), or rather affixoid replacement in the case of (16)b.

(16)   a.   fondamental*ismo* islamico   →   fondamental*ista* islamico
            'Islamic fundamentalism'   'Islamic fundamentalist'
       b.   bio*logia* molecolare   →   bio*logo* molecolare
            'molecular biology'   'molecular biologist'

Thus, a *fondamentalista islamico* is not a fundamentalist who is Islamic, but is someone who acts according to Islamic fundamentalism. This example of course reminds us of the cases discussed in (3), the difference being that in (16)a the *-ism / -ist* words are found in larger expressions.

Within the framework of CxM we can account for these data by resorting to two tools. The first is the notion of "phrasal lexeme" (Masini 2009) or, more appropriately

here, "phrasal names" (Booij 2009), i.e. fixed phrases that are encoded as lexical constructions. This is a crucial step of the whole reasoning, since it allows us to treat all the expressions in (15) and (16) as lexical units. The second is, obviously, second order schemas, which can be used to explicitly state a paradigmatic relationship (within the constructicon) between the two constructions in order to account for their production:

(17) $< [N_x A_i]_{Nj} \leftrightarrow [N \text{ with } SEM_x \text{ that has the property } SEM_i]_j> \approx$
      $<[[N_x+ista]_{Ny} A_i]_{Nz} \leftrightarrow [N \text{ with } SEM_y \text{ that has to do with } SEM_j]_z >$

The formula in (17) accounts for the set of data in (15), where the suffix -*ista* is attached to the head noun on the left. As for the examples in (16)a, we can adjust the formula already proposed in (4) above (see (18)), whereas another second order schema can be formulated along these lines to account for examples such as (16)b (see (19)).

(18) $< [[x\text{-}ismo]_{Ni} A_w]_{Nk} \leftrightarrow [SEM_i \text{ with the property } SEM_w]_k > \approx$
        $< [[x\text{-}ista]_{Nj} A_w]_{Nt} \leftrightarrow [\text{person with property Y related to } SEM_k]_t >$

(19) $< [[x\text{-}logia]_{Ni} A_w]_{Nk} \leftrightarrow [SEM_i \text{ with the property } SEM_w]_k > \approx$
        $< [[x\text{-}logo]_{Nj} A_w]_{Nt} \leftrightarrow [\text{person who is an expert of } SEM_k]_t >$

To conclude this section, it is worth pointing out that NA phrasal nouns are not the only phrasal lexemes to generate bracketing paradoxes of this kind. Also phrasal nouns of the NPN type – which are very common structures in Italian and Romance languages in general (Masini 2009) – act similarly, as illustrated below:

(20) a. pizza al taglio   → pizz*eria* al taglio
          'pizza sold by the slice'   'pizzeria where pizza is served by the slice'
      b. tennis da tavolo   → tenn*ista* da tavolo
          'table tennis'   'table tennis player'
      c. strumento a corde   → strument*ista* a corde
          'stringed instrument'   'strings player'

A *pizzeria al taglio* is not a *pizzeria* sold by the slice, but a place where pizza by the slice is sold (see also *pizza da asporto* 'takeaway pizza' → *pizzeria da asporto* 'takeaway pizzeria'); a *tennista da tavolo* is someone who plays table tennis and not a tennis player who has something to do with a table; and, finally, if I am a *strumentista a corde* I am not "stringed" in any way, but I play stringed instruments. Again, this form-meaning mismatch can be represented by means of a second order schema:

(21) $<[N_x P N_i]_{Nj} \leftrightarrow [N \text{ with } SEM_x \text{ related to } SEM_i]_j> \approx$
      $<[[N_x+SUFF]_{Ny} P N_i]_{Nz} \leftrightarrow [N \text{ with } SEM_y \text{ that has to do with } SEM_j]_z >$

        To sum up, in this section we showed that CxM is successful in explaining bracketing paradoxes in Italian that involve phrasal lexemes for two reasons. First, it recognizes phrasal nouns as constructions endowed with their own representation in the lexicon, so that it is possible to refer to their semantics as a whole. Second, the tool of

second order schemas guarantees that the correct semantics is computed out of the phrasal nouns, despite the form-meaning mismatch.

4. "Squeezed" phrasal lexemes in Russian

Russian displays a number of non-isomorphic phenomena that provide further evidence for the usefulness of second order schemas.

The first case we intend to discuss reminds us of the Bare Stem Constraint mentioned in Section 2 for Modern Greek (see (1)). Russian presents a large number of fixed expressions that consist of a (usually relational) adjective and a head noun (Masini & Benigni 2012: 420):

(22) a. mobil'nyj telefon
          mobile telephone
          'mobile phone'
      b. detskij sad
          child$_{ADJ}$ garden
          'kindergarten'

Similarly to what we proposed for Italian in Section 3, we consider these expressions as constructions (i.e. phrasal lexemes).[3] What is interesting is that some of these phrasal lexemes may function as a base for derivative processes: in (23) a relational adjective is formed (*železnodorožnyj*) which is formally and semantically linked to the phrasal noun *železnaja doroga* (Masini & Benigni 2012: 440):

(23) železnaja doroga   →   železn-o-dorož-nyj
      iron$_{ADJ}$ road   iron$_{STEM}$-LV-way$_{STEM}$-SUFF
      'railway'   'related to railway'

Like in Modern Greek, the phrasal noun undergoes some changes: the adjective becomes a stem, a linking vowel emerges, and a derivational suffix is added to the right of the noun *doroga*. Hence, the phrasal noun is somehow "compressed", it loses some of its constituting parts and new (typically morphological) elements intervene, thus giving rise, once again, to an at least partial mismatch.[4]

A more striking example of mismatch comes from so-called "stump compounds" (Comrie & Stone 1978), i.e. typically Russian (Soviet) expressions that are formed by combining an abbreviated form of a word with either a full word (24) or another abbreviated form (25) (what Billings 1998 calls "clips"):

---

[3] These lexical constructions differ in a number of ways from proper compounds in Russian. For details see Benigni & Masini (2009) and Masini & Benigni (2012).

[4] An extreme version of the Bare Stem Constraint occurs in cases such as Italian *cerchiobottismo*, noted by Gaeta (2003). *Cerchio-bott-ismo* (lit. ring-barrel-ism) 'trying to keep in with both sides' is a noun derived from the complex idiomatic expression *dare un colpo al cerchio e uno alla botte* (lit. to give a hit to the ring and one to the barrel) 'to run with the hare and hunt with the hounds'.

(24) a. *gor*odskoij   *sovet*   →   gorsovet
        city$_{Adj}$ council   'city council'
      b. *zar*abotnaja   *plata*   →   zarplata
        working   wage   'salary'
(25) *soc*ial'naja   *strach*ovka   →   socstrach
      social   insurance   'social insurance'

Given output forms such as *zarplata* or *socstrach*, it is difficult, if not impossible, to compute the correct semantics, since we do not have enough elements at the syntagmatic level. As pointed out by Masini & Benigni (2012: 430), stump compounds may be seen as a strategy to "squeeze" a pre-existing phrasal noun into a morphological, word-level unit. Indeed, in most cases, stump compounds are not derived from free phrases, but from established phrasal lexemes, especially of the AN type. The resulting stump compound has the same propositional content of the phrasal lexeme, but is typically connotated as a specialized term.

      There is still another case in Russian in which AN phrasal lexemes undergo some sort of "squeezing" and give rise to new complex words: so-called shortenings with the suffix *-ka*.[5] The bound form *-ka* is productively used as a "regular" diminutive suffix in Russian, however it is also found in examples such as the following (Masini & Benigni 2012):

(26) a. mineral'naja   voda   →   mineral-*ka*
        mineral   water
        'mineral water'
      b. maršrutnoe   taksi   →   maršrut-*ka*
        route$_{ADJ}$ taxi
        '(fixed) route taxi'
      c. èlektronnaja   počta   →   èlektron-*ka*
        electronic   mail
        'e-mail'

The data above are obtained by three steps: deletion of the head noun (*èlektronnaja počta* > *èlektronnaja*), truncation of the adjective, normally up to the stressed syllable (*èlektronnaja* > *èlektron*); and finally addition of *-ka* (*èlektron-ka*). The complex word *èlektronka* has the same propositional meaning of the phrasal noun *èlektronnaja počta*, but is characterized by a more familiar, intimate register.

      Obviously, the semantics of *èlektronka* cannot be computed by combining the meaning of its costituents. Exactly like in the two cases discussed above – relational adjectives derived from phrasal nouns on the one hand and stump compounds on the other –, we need to refer directly to the source construction, i.e. to the underlying phrasal noun, if we want to end up with the correct semantics. Once again, a possible solution is provided by second order schemas, which allow to make this role of the underlying phrasal noun explicit in the constructicon.

---

[5] *-ka* is not the only suffix used in these kinds of shortenings, but it is definitely the most common.

We propose a general formula for stump compounds derived from AN phrasal nouns in (27) and another for *-ka* formations in (28).[6]

(27) $< [A_i N]_{Nj} \leftrightarrow [N$ with the property $SEM_i]_j> \approx$
$< [A_{Stump}\text{-}N_{(Stump)}]_{Nz} \leftrightarrow [SEM_j [+specialized]]_z >$

(28) $< [A_i N]_{Nj} \leftrightarrow [N$ with the property $SEM_i]_j> \approx$
$< [A_{Trun}\text{-}ka]_{Nz} \leftrightarrow [SEM_j [+familiar]]_z >$

As we can see, the second schema directly refers to the semantics of the whole phrasal noun (SEM_j), represented in the first schema. In a certain way, we can say that second order schemas actually "save" the notion of compositionality, if we accept to define compositionality in a broader (non strictly syntagmatic) sense.

5. Nominalization of particle verbs in Dutch

The nominalization of Dutch particle verbs forms an intriguing challenge for linguistic analysis, and reveals how phrasal and morphological constructs may be related in a systematic fashion in the constructicon of a language. Particle verbs are complex predicates consisting of a particle and a verb, in most cases underived. They have the formal structure of either small verb phrases, or syntactic compounds. In other words, they are not morphological compounds, although they are lexical items (Booij 2010; Los et al. 2012). Quite often, the meaning of a particle verb is idiosyncratic, that is, its meaning cannot be derived from the meaning of its constituents.

The default nominalization in Dutch is that by means of the suffix *-ing*, as illustrated by the following examples of a simplex verb, a prefixed verb, a suffixed verb and a particle verb respectively:

(29) *verb*   *deverbal noun*
boek 'to book'   boek-ing 'booking'
aan-bíd 'to worship'   aanbidd-ing 'worship'
central-iseer 'to centralize'   centraliser-ing 'centralization'
áan-bied 'to offer'   aanbied-ing 'offer'

In addition to this default process, there are a number of (unproductive, in present-day Dutch) nominalization processes, that have applied in particular to simplex verbs of German origin:

(30) *verb*   *deverbal noun*
val 'fall'   val 'fall'
bind 'bind'   band 'bond'
bied 'offer'   bod 'offer'

---

[6] Please note that this is a simplified formalization, where the phonetic form of the stump constituents and of the *-ka* forms are just hinted at (A_{Stump}, $N_{(Stump)}$, A_{Trun}) and not detailed. The abbreviated form of N in (27) is optional ($N_{(Stump)}$) so as to cover both stump compounds proper (24) and clips (25).

kom 'come'	kom-st 'coming'
geef 'give'	gav-e 'gift, talent' / gif-t 'gift'

In some cases, the existence of an unproductive type blocks the application of the default process, in other cases the simplex verb has more than one nominalization, with a corresponding semantic difference:

(31)   *verb*   *deverbal noun*
       breek 'break'   brek-ing 'refraction'
                                 breuk 'fracture'
       bied 'offer   bied-ing 'bidding'
                                 bod 'offer'
       but
       val 'fall'   *val-ling / val
       geef 'give'   *gev-ing / gif-t
       kom 'come'   *kom-ing / kom-st

However, words like *geving* and *koming* do occur as parts of complex words such as *bericht-geving* 'lit. report-giving, reporting' and *tegemoet-koming* 'lit. towards-coming, concession'. This is a case of 'embedded productivity' (Booij 2010): a word formation process may be applied productively to a base word in larger morphological structures, even though it cannot be applied productively to that base word in isolation. A more detailed study of these facts can be found in Booij (2014).

Our concern in this article is the fact that, quite often, Dutch particle verbs have the same unproductive form of nominalization as their corresponding verbal base, as illustrated by the following examples:

(32)   *verb*   *nominalization*
       a.   bied 'to offer'   bod 'offer'
            aan-bied 'to offer'   aan-bod 'offer'
       b.   val 'to fall'   val 'fall'
            aan-val 'to attack'   aan-val 'attack'
       c.   kom 'to come'   kom-st 'arrival'
            aan-kom 'to arrive'   aan-kom-st 'arrival'
       d.   slaan 'to hot'   slag 'hit'
            op-slaan 'to store'   op-slag 'storage'
       e.   geef 'to give'   gav-e 'gift'
            uit-geef 'to publish'   uit-gav-e 'publication'
       f.   zien 'to see'   zicht 'sight'
            toe-zien 'to supervise'   toe-zicht 'supervision'

These data illustrate that the deverbal nominalization of simplex verbs in Dutch is lexically governed to a high degree. The crucial observation is that the nominalized form of a particle verb may consist of the particle followed by the deverbal noun of the corresponding simplex verb. If the particle verb nominalizations were derived in a straightforward syntagmatic fashion, one would expect that they would all be created by

means of suffixation with *-ing*, as this is the default suffix for complex verbs. However, this is not the case, and instead the lexically governed nominalized form of the verbal part of the particle verb is often used as a building block of the nominalization of these particle verbs. Note that the default nominalization with *-ing* can apply to particle verbs, as illustrated by the noun *aanbieding* 'offer' (besides *aanbod*) mentioned above. Default nominalization always applies to those particle verbs in which a particle combines with a noun or an adjective, thus triggering conversion of that noun or adjective into a verb:

(33) *base word*   *particle verb*   *derived noun*
      hok 'cage'   op-hok 'encage'   ophokk-ing 'encagement'
      tak 'branch'   af-tak 'branch'   aftakk-ing 'branch'

      slank 'lean'   af-slank 'slim'   afslank-ing 'slimming'
      zwak 'weak'   af-zwak 'weaken'   afzwakk-ing 'weakening'

That is, only particle verbs whose base verb has an unproductive nominalization type allows for an unproductive type of nominalization, and only the same as its base verb.

This generalization can be accounted for by assuming that nominalizations of particle verbs of unproductive types consist of a particle + the nominalized form (V-Nom) of the simplex verb:

(34)   [Part [V-Nom]$_N$]$_N$

The form of Nom is variable: zero (conversion), stem allomorphy (as in *slag, zicht*), or a suffix (*-e, -t* or *-st*). Structure (34) correctly predicts that the nominalization type of a particle verb normally corresponds to the nominalized type of the corresponding simplex verb. This means that from a formal point of view such nominalizations are nominal compounds of which the head is deverbal. However, from a semantic point of view, they are to be interpreted as nominalizations of particle verbs.

There is independent evidence that the compound schema (34) is needed for Dutch, as there are many nominal compounds with a particle in non-head position. A telling example is the class of compounds that begin with the particle *toe*. This word is not used as an adverb or adposition in isolation. We find it only as part of the discontinuous adposition *naar ... toe*, in particle verbs, and in compounds:

(35) a.   naar Amsterdam toe
           to Amsterdam to 'to Amsterdam'
      b.   toe-stuur (V)
           to–send 'send to'
      c.   toe-gang (N)
           to-going 'access'
      d.   toe-loop (N)
           to-walk 'run-up'

The two *toe*-nouns in (35) do not have a corresponding particle verb **toegaan* or **toelopen*.

14

There are many other compounds of this type for which the corresponding particle verb does not exist at all, or does exist but not with the relevant meaning, for instance:

(36)   af-komst 'descendance/ *af-komen 'to descend' (af-komen 'to come off')
        af-val 'waste' / *af-vallen 'to waste' (af-vallen 'to loose weight')
        af-stand 'distance' / * af-staan 'to distance' (af-staan 'to give')

What we should be able to express is the following generalization: a compound noun of this type (particle + deverbal noun) is interpreted as the nominalization of the corresponding particle verb, if that particle verb exists.

Let us illustrate this by listing the nominalized form of a number of particle verbs of which the verbal constituent features one of these unproductive nominalization types:

(37)   *verb*   *deverbal noun*
        grijp 'grab'   greep 'grip'
        in-grijp 'intervene'   in-greep 'intervention'
        mis-grijp 'slip up'   mis-greep 'slip-up'

        zien 'see'   zicht 'sight'
        aan-zien 'watch'   aan-zicht 'view'
        in-zien 'understand'   in-zicht 'insight'
        toe-zien 'supervise'   toe-zicht 'supervision'
        op-zien 'supervise'   op-zicht 'supervision'

        gaan 'go'   gang 'going'
        af-gaan 'go down'   af-gang 'flop'
        door-gaan 'go on'   door-gang 'thoroughfare'
        in-gaan 'enter'   in-gang 'entrance'
        neer-gaan 'go down'   neer-gang 'downturn'
        op-gaan 'rise'   op-gang 'ascent'
        voort-gaan 'continue'   voort-gang 'continuation'

        staan 'stand'   stand 'stand'
        achter-staan 'lag behind'   achter-stand 'lag'
        af-staan 'abandon'   af-stand 'abandonment'
        op-staan 'rise'   op-stand 'rising'

        slaan 'hit'   slag 'hit'
        aan-slaan 'strike'   aan-slag 'attack'
        af-slaan 'sell by auction'   af-slag 'auction'
        door-slaan 'hit'   door-slag 'hit'
        in-slaan 'smash'   in-slag 'smash'
        mis-slaan 'miss'   mis-slag 'miss'
        op-slaan 'store'   op-slag 'storage'
        over-slaan 'pass over'   over-slag 'pass-over'

In the case of the verb *nemen* 'to take', the nominalized form name does not appear in isolation. However, it is used in complex morphological structures, such as *in-bezit-name* 'taking possession of', a case of embedded productivity.

(38)   *verb*   *deverbal noun*
       neem 'take'   *name (but in-bezit-name 'inpossession-taking')
       af-neem 'decrease'   af-name 'decrease'
       in-neem 'take in   in-name 'intake'
       op-neem 'record'   op-name 'recording'
       over-neem 'take over'   over-name 'take-over'
       toe-neem 'increase'   toe-name 'increase']

The observed mismatch between the form and the meaning of these nominalizations of particle verbs can now be accounted for by the following second order schema:

(39) $< [Part_i \; V_j]_k \leftrightarrow SEM_k > \; \approx \; < [Part_i \; [V_j\text{-}Nom]_N]_{Nm} \leftrightarrow [NOM \; [SEM_k]]_m >$

The semantic operator NOM stands for the semantic effect of nominalization. SEM_k stands for the meaning of the particle verb as a whole, and since NOM has scope over SEM_k, it is expressed that compounds consisting of a particle and a deverbal noun are interpreted as nominalizations of the corresponding particle verbs. Thus, we see another argument here for the necessity of second order schemas in the lexicon.

In this analysis, we interpret nominalizations of particle verbs such as *aankomst* 'arrival' as nominal compounds, since this is the only morphological structure available for combining a particle and a deverbal noun. We do not have 'separable complex nouns' besides separable complex verbs, and hence, when combined with a noun, a particle can only appear as the left constituent of a nominal compound. This interpretation is in accordance with the stress patterns of these nouns, as they have main stress on their first constituent, just like nominal compounds in general.

6. Conclusions

In this paper we tackled the problem of compositionality in complex words. We provided data from various languages that show that a strict definition of compositionality such as the one formulated in Section 1 does not hold for a number of cases. We showed that there are many word formation processes in various languages that display a mismatch between form and meaning. In order to overcome this generalized problem, we proposed to use second order schemas within the framework of CxM, i.e. paradigmatic relations among word formation schemas to be explicitly encoded in the constructicon. In this way, the need for a complete isomorphism between the constituents of a complex word and its resulting meaning vanishes, since also the relevant constructions, as well as the paradigmatic relations between them, play an active role in computing the final meaning.

References

Aronoff, Mark. 1976. *Word formation in generative grammar*. Cambridge, Mass.: The MIT Press.

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*, ed. Geert Booij and Jaap van Marle, 1-26. Dordrecht: Kluwer.

Benigni, Valentina and Francesca Masini. 2009. Compounds in Russian. *Lingue e Linguaggio* 8(2): 171-193.

Billings, Loren. 1998. *Morphology and syntax: delimiting stump compounds in Russian*. In *Proceedings of the First Mediterranean Morphology Meeting*, ed. Geert Booij, Angela Ralli and Sergio Scalise, 99-110. Patras: University of Patras.

Bisetto, Antonietta and Francesca Moschin. 2010. Bracketing paradoxes as constructions. Paper presented at the 14th International Morphology Meeting (IMM14), Budapest, May 13-16, 2010. http://www.nytud.hu/imm14/abs/bisetto_moschin.pdf (abstract accessed 28 May 2014).

Bisetto, Antonietta and Sergio Scalise. 1991. Compounding: Morphology and/or syntax? In *The boundaries of morphology and syntax*, ed. Lunella Mereu, 31-48. Amsterdam / Philadelphia: John Benjamins.

Booij, Geert. 1997a. Allomorphy and the autonomy of morphology. *Folia Linguistica* 31(1-2), 25-56.

Booij, Geert. 1997b. Autonomous morphology and paradigmatic relations. In *Yearbook of Morphology 1996*, ed. Geert Booij and Jaap van Marle, 35-53. Dordrecht: Kluwer.

Booij, Geert. 2002. *The morphology of Dutch*. Oxford: Oxford University Press.

Booij, Geert. 2009. Phrasal names: A constructionist analysis. *Word Structure* 2(2): 219-240.

Booij, Geert. 2010. *Construction morphology*. Oxford: Oxford University Press.

Booij, Geert. 2014. Nominalization of particle verbs. Ms.

Botha, Rudolf. 1984. *Morphological mechanisms: Lexicalist analyses of synthetic compounding*. Oxford: Pergamon Press.

Comrie, Bernard and Gerald Stone. 1978. *The Russian language since the revolution*. Oxford: Clarendon Press.

DiSciullo, Anna Maria and Edwin Williams. 1987. *On the definition of word*. Cambridge, Mass.: The MIT Press.

Dowty, David. 2007. Compositionality as an empirical problem. In *Direct compositionality*, ed. by Chris Barker & Pauline Jacobson, 23-101. Oxford: Oxford University Press.

Frege, Gottlob. 1892. Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik* 100(1), 25-50.

Gaeta, Livio (2003). Ai limiti della morfologia basata sulle parole. In *Scritti di morfologia in onore di Sergio Scalise in occasione del suo 60° compleanno*, ed. Antonietta Bisetto, Claudio Iacobini and Anna M. Thornton, 47-59. Roma: Caissa Italia.

Ghomeshi, Jila, Ray Jackendoff, Nicole Rosen and Kevin Russell. 2004. Contrastive

focus reduplication in English (The salad-salad paper). *Natural Language and Linguistic Theory* 22(2): 307-357.

Goldberg, Adele. 1995. *Constructions. A Construction Grammar approach to argument structure*. Chicago: The University of Chicago Press.

Goldberg, Adele. 2006. *Constructions at work. The nature of generalization in language*. Oxford: Oxford University Press.

Hoeksema, Jack. 2012. Elative compounds in Dutch: Properties and developments. In *Intensivierungskonzepte bei Adjektiven und Adverben im Sprachenvergleich / Crosslinguistic comparison of intensified adjectives and adverbs*, ed. Guido Oebel, 97-142. Hamburg: Verlag dr. Kovač.

Jackendoff, Ray. 2013. Constructions in the Parallel Architecture. In *The Oxford Handbook of Construction Grammar*, ed. Thomas Hoffmann and Graeme Trousdale, 70-92. Oxford: Oxford University Press.

Kapatsinski, Vsevolod. 2013. Conspiring to mean: Experimental and computational evidence for a usage-based harmonic approach to morphophonology. *Language* 89(1): 110-148.

Los, Bettelou, Corrien Blom, Geert Booij, Marion Elenbaas and Ans Van Kemenade. 2012. *Morphosyntactic change: A comparative study of particles and prefixes*. Cambridge: Cambridge University Press.

Masini, Francesca and Valentina Benigni. 2012. Phrasal lexemes and shortening strategies in Russian: The case for constructions. *Morphology* 22(3): 417-451.

Masini, Francesca. 2009. Phrasal lexemes, compounds and phrases: A constructionist perspective. *Word Structure* 2(2), 254-271.

Morris, Caroline. 2013. *Bekje-af, knettertjegek and dolletjesgelukkig. The use and development of intensifying diminutive compounds in Dutch within the framework of Construction Morphology*. Groningen: University of Groningen MA Thesis.

Namer, Fiammetta. 2013. Adjectival bases of French *-aliser* and *-ariser* verbs: Syncretism or under-specification? In *Morphology in Toulouse. Selected Papers from the 7th Décembrettes (Toulouse, 2-3 December 2010)*, ed. Nabil Hathout, Fabio Montermini and Jesse Tcheng, 185-210. München: LINCOM.

Nesset, Tore. 2008. *Abstract phonology in a concrete model: Cognitive linguistics and the morphology-phonology interface*. Berlin / New York: Mouton de Gruyter.

Ralli, Angela. 2013. *Compounding in Modern Greek*. Dordrecht: Springer.

Spencer, Andrew. 1988. Bracketing paradoxes and the English lexicon. *Language* 64(4): 663-682.

Spencer, Andrew. 1991. *Morphological theory: An introduction to word structure in Generative Grammar*. Oxford: Blackwell.

Vallès, Teresa. 2003. Lexical creativity and the organization of the lexicon. *Annual Review of Cognitive Linguistics* 1: 137-160.

Virgillito, Daniele. 2010. *Bracketing paradoxes in Italian*. Bologna: University of Bologna MA Thesis.

Williams, Edwin. 1981. On the notions 'lexically related' and 'head of a word'. *Linguistic Inquiry* 12(2): 245-274.