

A Taxonomy for Afrikaans and Dutch Compounds

Gerhard B van Huyssteen

Centre for Text Technology

North-West University

Potchefstroom, South Africa

gerhard.vanhuissteen@nwu.ac.za

Ben Verhoeven

CLiPS - Computational Linguistics

University of Antwerp

Antwerp, Belgium

ben.verhoeven@uantwerpen.be

Abstract

The linguistic categorisation of compounds dates back to some of the earliest work in linguistics. The cross-linguistic compound taxonomy of Bisetto and Scalise (2005), later refined in Scalise and Bisetto (2009), is well-known in linguistics for understanding the grammatical relations in compounds. Although this taxonomy has not been used extensively in the field of computational linguistics, it has the potential to influence choices with regard to compound annotation and understanding in natural language processing. For example, their 2005 taxonomy formed the basis for the large-scale, multilingual database of compounds, called CompoNet. The aim of this paper is to examine their latest taxonomy critically, especially with a view on rigorous implementation in computational environments (e.g. for the morphological annotation of compounds). We propose a number of general improvements of their taxonomy, as well as some language-specific refinements.

1 Introduction

The CompoNet database¹ is a large database of compounds from 27 different languages, which was developed at the Department of Foreign Languages of the University of Bologna, in collaboration with native speaker linguists. The database can be used to study compounding of a given language, of a given family (e.g. Germanic, Slavic, etc.), and compounding in general from a typological perspective. Fields in the database include, inter alia, the compound and its part-of-speech (POS) category; the components in the compound and their respective POS categories; the structure of the compound (e.g. [N+N]); whether it is endocentric or exocentric, and an indication of the position of the categorial and semantic head; some inflectional information (plural and gender); glosses; and the classification category of the compound.

With regard to the latter, the well-known classification taxonomy of Bisetto and Scalise (2005) is used (see Figure 1). This classification scheme is based on the view that the grammatical relations between the components of a compound are similar to those in syntactic constructions, *viz.* subordinate, attributive, and coordinate relations. In addition, each of these types can be endocentric or exocentric, depending of the presence (endocentric) or absence (exocentric) of a head constituent.

In a project on automatic compound processing (the AuCoPro project; see <http://tinyurl.com/aucoopro>), we investigated various aspects related to the computational processing of compounds (Verhoeven et al., 2014). In a specific subpart of this project, we aimed to gain more insight in compound semantics in general by drawing from perspectives from computational semantics (i.e. Ó Séaghdha, 2008), typological studies (e.g. Lieber, 2009a; Scalise & Bisetto, 2009), and

This work is licensed under a Creative Commons Attribution 4.0 International Licence. Page numbers and proceedings footer are added by the organisers. Licence details: <http://creativecommons.org/licenses/by/4.0/>

¹ <http://componet.sslmit.unibo.it/>

construction-based approaches to word-formation (i.e. cognitive grammar (Langacker, 2008) and construction morphology (Booij, 2010)). In addition, we specifically wanted to add Afrikaans compounds to the CompoNet database (as Afrikaans was not included in the original CompoNet project), as well as revise the existing Dutch compounds in CompoNet (based on the insights of the AuCoPro project). As a first phase, we made 56 changes to the Dutch database (mostly correcting minor spelling and classification errors, as well as adding a few additional, prototypical examples), and added 144 Afrikaans compounds to the database (compared to a total of 188 Dutch compounds; the 144 Afrikaans compounds were representative of all part-of-speech categories that can be found in Afrikaans compounds).

However, soon after the project commenced, we encountered some limitations with the original CompoNet annotation guidelines, specifically with regard to the classification of compounds. In Section 2 we give an overview of these problems, and discuss some recent literature on the classification of compounds. In Section 3 we describe our solution to these limitations by postulating a classification scheme that would be suitable for rigorous implementation in computational environments (e.g. for the morphological annotation of compounds). We conclude this paper with a discussion of future research.

2 Previous work

In a publication of this nature, it is impossible to discuss all previous research, or even the details of some of the literature influencing our own taxonomy for Afrikaans and Dutch (see Section 3); suffice to point to the overview and summary provided by Scalise and Bisetto (2009), as well as applications of their framework by Lieber (2009a, 2009b). In the remainder of this section we therefore only focus on those aspects that influenced our own taxonomy.

During the initial phase of the project, we encountered a number of stumbling-blocks with regard to the annotation guidelines. As indicated above, compound classification in the CompoNet database is based on Bisetto and Scalise (2005) (see Figure 1). However, since then, Scalise and Bisetto (2009) have revised their original taxonomy (see Figure 2), and the dilemma was therefore that we could not take cognisance of these new insights (e.g. the distinction between root and verbal-nexus compounds, or between attributives and appositives), since we had to stay as close as possible to the original annotation guidelines for purposes of cross-lingual compatibility. Table 1 provides a summary of some of the most important notions in Scalise and Bisetto’s (2009) taxonomy, some additional remarks by Lieber (2009a, 2009b), and examples provided by them. Other summaries of their framework include Arcodia et al. (2009); Arnaud and Renner (2014); Vercellotti and Mortensen (2012).

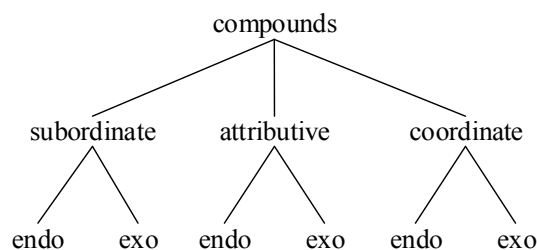


Figure 1. Compound taxonomy of Bisetto and Scalise (2005)

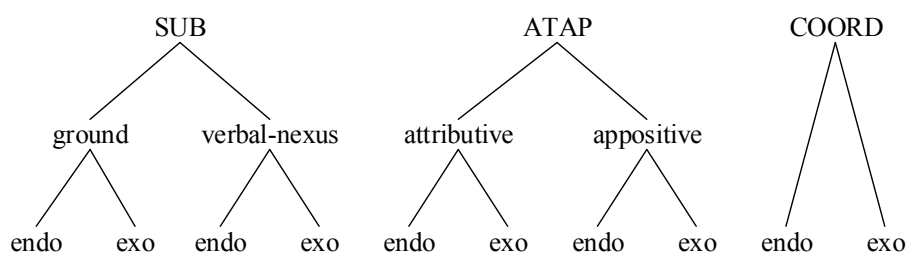


Figure 2. Compound taxonomy of Scalise and Bisetto (2009)

| Concept | Key definitional aspects | Examples |
|--------------------------|--|--|
| Subordinate | <ul style="list-style-type: none"> Components share a head-complement relation (subordination) Argumental relation between components (Lieber, 2009a: 93) At least one of the features of the head constituent is to match the encyclopaedic features that characterise the non-head Includes synthetic compounds (Lieber, 2009b: 359), and neoclassical compounds Among the most widely attested of compound types (specifically endocentric; Lieber, 2009a: 93) | See below under “Ground subordinate” and “Verbal-nexus subordinate” |
| Ground subordinate | <ul style="list-style-type: none"> Corresponds to root/primary compounds Lexemes can be both simple and complex When complex and includes a verb, it is incapable of influencing the interpretation of the compound [no examples provided] Semantic relation between constituents is influenced by semantico-encyclopaedic information NN compounds with an ‘of’ relation (Lieber, 2009a: 88), but also if they have a (quasi-)argumental relation (e.g. <i>cookbook author</i>) (Lieber, 2009b: 359) | <i>windmill</i> (endocentric) <i>mushroom soup</i> (endocentric) <i>love story</i> (endocentric) <i>steam boat</i> (endocentric) <i>coffee cup</i> (endocentric) |
| Verbal-nexus subordinate | <ul style="list-style-type: none"> Corresponds to secondary/syntactic compounds Presence of verb (or any other deverbal constituent) as head Verbs select the non-head semantically, be it an argument (<i>bookseller</i>) or a complement/adjunct (<i>street seller</i>) Quintessential example is synthetic compound (Lieber, 2009a: 88) | <i>truck driver</i> (endocentric) <i>cost containment</i> (endocentric) <i>city employee</i> (endocentric) <i>pickpocket</i> (exocentric) <i>killjoy</i> (exocentric) <i>cut-throat</i> (exocentric) |
| Attributive | <ul style="list-style-type: none"> Non-head (often an adjective) expresses a quality of the head [often a noun] (i.e. head is modified by a non-head expressing a ‘property’ of the head) The non-head fulfils at least one of the encyclopaedic features of the head; it has an ‘adjectival’ function Clear argumental relationship between constituents lacks (Lieber, 2009b: 359) Default semantic type (Lieber, 2009a: 97) Most frequently attested in the languages of the world (Lieber, 2009a: 97) | <i>high-school</i> (endocentric) <i>blue-eyed</i> (endocentric) <i>blue cheese</i> (endocentric) <i>atomic bomb</i> (endocentric) <i>redskin</i> (exocentric) <i>greenhouse</i> (exocentric) <i>freelance</i> (exocentric) |
| Appositive | <ul style="list-style-type: none"> Non-head expresses a property of the head by means of a noun acting as an attribute Noun plays an attributive role and is often interpreted metaphorically Non-head can also be a verb [when the head is an adjective] NN compounds cannot be paraphrased with ‘of’ (Lieber, 2009a: 88) | <i>snail mail</i> (endocentric) <i>swordfish</i> (endocentric) <i>mushroom cloud</i> (endocentric) Du. <i>druipnat</i> (endocentric) |
| Coordinate | <ul style="list-style-type: none"> Constituents with an ‘and’ relation Two semantic heads, but only one act as categorial head Could be additive (<i>Baden-Württemberg</i>), or redundant (<i>palm tree</i>) Coordinates could be, <i>inter alia</i>, reduplicates (It. <i>lecca-lecca</i> ‘lolly-pop’) | <i>bittersweet</i> (endocentric) <i>poet-doctor</i> (endocentric) <i>woman doctor</i> (endocentric) <i>Austria-Hungary</i> (exocentric) <i>mother-child</i> (exocentric) <i>north-east</i> (exocentric) |

Table 1. Verbatim summary of Scalise and Bisetto (2009) with additional remarks by Lieber (2009a, 2009b), and our remarks in square brackets

There are two significant differences between these two taxonomies: the label ATAP (ATtributive-APpositive) is introduced in the 2009 version; and a new categorisation level is introduced in the 2009 version to make a distinction between root and verbal-nexus compounds. With the introduction of the “artificial” ATAP label (placed on the same hierarchical level as subordinate and coordinate compounds) as a superordinate category for attributive and appositive compounds, Scalise and Bisetto (2009) lost some “correctness”. In the new taxonomy attributives and appositives are therefore on the same categorisation level as verbal-nexus and ground compounds, which, in our opinion, is incorrect. (Fábregas and Scalise (2012) later replace attributive compounds to its original hierarchical level, and then distinguish between two types of attributive compounds, *viz.* true attributives, and appositives. Also see Arnaud and Renner (2014), and Vercellotti and Mortensen (2012: 572) for a critique of the categorisation levels used by Scalise and Bisetto (2009).)

The only annotation protocol available for CompoNet is the article by Bisetto and Scalise (2005), as well as some notes for some of the languages on the CompoNet website (only available to registered users). In our experience, these guidelines were not always explicit or elaborate enough (see also Vercellotti and Mortensen, 2012: 547), and in addition, were sometimes difficult to interpret given other discussions in the literature (notably Fábregas and Scalise, 2012; Lieber, 2009a, 2009b; Scalise and Bisetto, 2009). Two examples suffice. Firstly, in Table 1 we indicated that Scalise and Bisetto (2009: 48-49) distinguish between subordinate and attributive compounds by the manner in which the head selects the non-head: in subordinate compounds “at least one of the features of the head constituent is **to match** the encyclopaedic features that characterise the non-head” (with *apple cake* as an example), while in attributive compounds “the non-head **fulfils** at least one of the encyclopaedic features of the head” (with *snail mail* as an example). In our opinion, this should be the other way round: in SNAIL and MAIL the property SLOW provides the **match** between the two constituents, whereas APPLE **fulfils** the INGREDIENT part of the concept CAKE.

A second example that confuses, comes from Lieber (2009a): on p. 98, with regard to *dog bed* as an example of an attributive compound, she states that “there is no verbal element here, so a subordinate interpretation is ruled out”. However, on p. 93 she lists *table leg* as one of the first examples of endocentric subordinate compounds, despite the fact that there is also no verbal element in *table leg*. Similarly, Vercellotti and Mortensen (2012: 549) interprets Scalise and Bisetto’s (2009) differentiation between verbal-nexus and ground compounds on the basis that the former have verb-argument/adjunct relations, while the latter have *no verbs*. However, Scalise and Bisetto (2009: 51) says about ground compounds containing complex lexemes: “when they *include a verb*, this is incapable of influencing the interpretation of the compound” (our emphasis). Although examples like these might be trivial (and does not take away anything from the overall insight in the categorisation of compounds), they do cause some confusion for the annotator who is provided with these publications as annotation guidelines.

Lastly, one of the problems we had with the original taxonomy was that it was not rich enough to allow for all compound types in Afrikaans and Dutch to be categorised, or at least not powerful enough to distinguish between various kinds of compounds. For example, in the original database a separable complex verb (SCV) like the Dutch (Du.) *neer+gooien* **down+throw** ‘to throw down’ was categorised as an attributive compound, while it should in reality rather be categorised as “Other” (OTH), a category in CompoNet reserved for examples that do not fit any of the other categories. Other examples include the difference between synthetic compounds (like the Afrikaans (Afr.) *gras+sny-er* **grass+cut-extN** ‘lawn mower’²) and parasyntetic compounds (Afr. *glad+maak-ing* **smooth+make-extN** ‘smoothing’), compounding compounds (Du. *oude+mannen+huis* **old+men+house** ‘retirement home for men’), and reduplications (Afr. *speel_-+speel* **play_LINK+play** ‘easily’). This illustrates that any taxonomy should at least provide for a slot for language-specific or other marginal phenomena – an aspect we will introduce in Section 3.

3 New proposal

In motivating why they came up with a revised taxonomy, Scalise and Bisetto (2009: 49) state that, given “the evolution of science, the need has arisen to add further levels of analysis to the classification”. They also invite further amendments to their newly proposed taxonomy, but warn that “anyone wanting to follow up on this issue will necessarily have to come to grips ... with the diverse compound formations that populate the languages of the world” (Scalise and Bisetto, 2009: 53). In as such, our new proposal wants to suggest some refinements to the general taxonomy of Scalise and Bisetto (2009) on the one hand, and on the other hand wants to make some language-specific changes pertaining to Afrikaans and Dutch (with the possibility that it could also be applicable to other (Germanic) languages). Furthermore, it should be kept in mind that our taxonomy also has a secondary aim, namely to serve as a structure for annotation of compounds in a database like CompoNet.

Our proposed taxonomy is presented in Figure 3, while Table 2 (as an Appendix) explicates this taxonomy with construction schemas for prototypical endocentric compounds, as well as an Afrikaans example for each instance. Although only Afrikaans examples are listed, we do claim that the

² Following the conventions in CompoNet, we use the following abbreviations: extN=nominaliser; extV=verbaliser; extAdj=adjectiviser; extAdv=adverbialiser; Sw=semi-word.

taxonomy holds true for Dutch: all categorial patterns listed by De Haas & Trommelen (1993) have been accounted for in some or other way in the taxonomy. In the remainder of this section, we explain and motivate only those aspects of our taxonomy that differ from Scalise and Bisetto (2009).

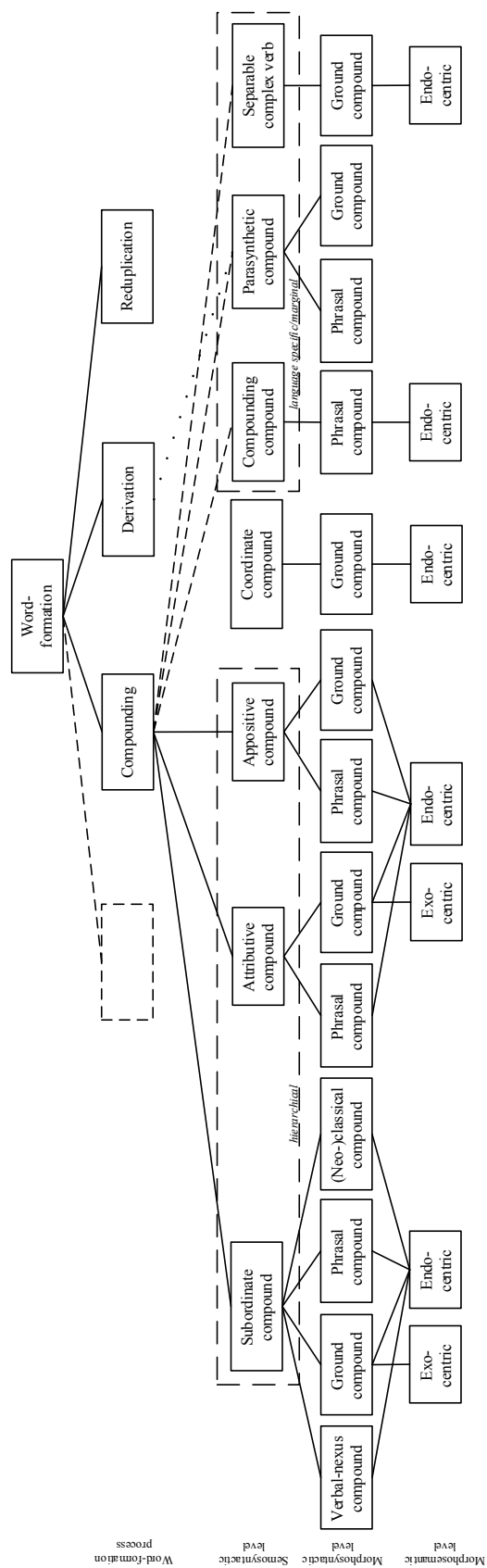


Figure 3. Taxonomy for Afrikaans and Dutch compounds (adapted from Van Huyssteen, 2014)

One of the important aspects of any taxonomy, is that taxa of the same type should be placed on the same taxonomic level/rank (e.g. *dog* and *cat* are on the same taxonomic rank), and that criteria for such ranks should be made explicit. In our taxonomy, we show that compounding, derivation, reduplication, etc., are all word-formation processes. We explicitly include reduplication here, in order to clarify that we do not consider examples like Afr. *dokter-dokter* **doctor-doctor** ‘play doctor’, or Du. *snel+snel* **quick+quick** ‘very quick’ to be compounds, as some authors like Kempen (1969) believe, but rather consider it a separate word-formation process (following Fábregas and Scalise, 2012, and Van Huyssteen, 2004). Note also the dotted line that links derivation with parasynthetic compounds, because parasynthetic compounds are formed through a derivational process: compounding by means of derivation (Booij and Van Santen, 1998: 178; see discussion below).

With regard to compounding specifically, we maintain the three taxonomic ranks of Scalise and Bisetto (2009) (unlike Vercellotti and Mortensen (2012) who added another level). On the semosyntactic level, the grammatical relations between constituents in compounds are used as classification criterion, and four types of relations are distinguished: subordinate, attributive, appositive, and coordinate. These four taxa operate on the same level of categorisation, and not like in the 2009-version of the Scalise and Bisetto taxonomy, where attributive and appositive were positioned on the same level as verbal-nexus and ground compounds. Following Vercellotti and Mortensen’s (2012) insight that subordinate, attributive and appositive compounds are more similar to each other than to coordinate compounds (i.e. the latter is not hierarchical), we lump them together with a dotted line in an area marked “hierarchical” (indicating their shared characteristic).

Note also that Vercellotti and Mortensen (2012) discard the notion of appositive compounds, since: (1) it “is unclear how many languages would need this category, given the difficulty distinguishing the category”; and (2) “‘appositive’ is already in the literature as a type of coordinate compound” (2012: 574). We have to agree to some degree with them on both accounts, but nonetheless maintain appositive as a useful label. Compare an appositive compound like Du. *sleutel+woord* **key+word** ‘keyword’ with a coordinate compound like Du. *dichter-zanger* **poet-singer** ‘idem’. A *sleutelwoord* is a word that is like a key, but nonetheless still a word; it is not a key that is also a word. In contrast, a *dichter-zanger* is a singer that happens to be a poet as well, but could just as well be paraphrased as a poet that happens to be a singer. Hence, we maintain that there is a difference between appositives and coordinates, with the former being right-headed, and the latter (at least semantically) dual-headed. Similarly, an appositive is subtly different from an attributive compound (and it is therefore often difficult to distinguish the two from each other; see also Arcodia et al. (2009), and Arnaud and Renner (2014)). A *sleutelwoord* is not a kind of word of the same order as a Du. *taboe+woord* **taboo+word** ‘taboo word’, or a Du. *mode+woord* **fashion+word** ‘trendy word’: a *sleutelwoord* is a word that **is like** a key, while a *taboewoord* is not like a taboo – the word **is** a taboo; a *modewoord* is not like the fashion, it **is** fashion. We maintain that appositives most often have an ‘is like’ metaphorical interpretation, while attributives have a literal ‘(that/which) is’ relation.

On the semosyntactic level, we can now formulate high-level construction schemas (Booij, 2010) for each of the four major endocentric compound types, as the bold parts in (1) to (4). To illustrate, read (1) as follows: on the phonological pole, a word $[a]_{xi}$ (*table*) can combine with another word $[b]_{xk}$ (*leg*) to form a new word $[ab]_{xk}$ (*table leg*), which as a whole ($_k$) should be interpreted on the semantic pole as $[LEG_j$ of $TABLE_i]_k$; note that $_i$, $_j$ and $_k$ are indices that mark the identity of constituents on the phonological and semantic poles (i.e. on the left and right of the double-arrow respectively).

- (1) Subordinate compounds: $[[a]_{xi} [b]_{xj}]_{xk} \leftrightarrow [SEM_j \text{ of } SEM_i]_k$
where the $_x$ of $[a]=N/V/Adj/Adv/Num/P/Phrase/Sw$; the $_x$ of $[b]=N/Adj/V/V-extN/V-extAdj/Sw^3$
- (2) Attributive compounds: $[[a]_{xi} [b]_{xj}]_{xk} \leftrightarrow [SEM_j \text{ is } SEM_i]_k$
where the $_x$ of $[a]=Adj/Adv/AP/Num/Phrase$; the $_x$ of $[b]=N/Adj$
- (3) Appositive compounds: $[[a]_{xi} [b]_{xj}]_{xk} \leftrightarrow [SEM_j \text{ like } SEM_i]_k$
where the $_x$ of $[a]=N/V/P/Phrase$; the $_x$ of $[b]=N/Adj$
- (4) Coordinate compounds: $[[a]_{xi} [b]_{xj}]_{xk} \leftrightarrow [SEM_i \text{ and/or } SEM_j]_k$
where $_x=N/V/Adj/Adv/P$

³ In Afrikaans, a pronoun can also act as head, as in the construction Afr. *ma-hulle* **mother-they** ‘mother and them’.

On the second taxonomic rank, the morphosyntactic level, compounds are distinguished in terms of the morphosyntactic (categorical) nature of the constituents, i.e. whether it is a lexical word, a phrase, or semi-word; in (1) to (4) these constituents are indicated in italics. All four types of compounds can be formed by means of ground words (i.e. uninflected words), which could be either simplex (e.g. *gebruik* in Afr. *gebruik+sfeer* **usage+sphere** ‘usage sphere’), or complex (e.g. Afr. *gebruik-er* **use-extN** ‘user’ in *gebruiker+vriendelik* **user+friendly** ‘user friendly’). All the major word categories can function as constituents in compounds, including N, V, Adj, Adv, Num, and P.

All except coordinate compounds can take phrasal elements as non-heads; these could range from full sentences (Afr. *Sannie-gaan-weeshuis-toe-rokkie* **Sannie-goes-orphanage-to-dress** ‘worn-out dress’), phrases (NP, VP, AP, PP), or phrase-like phrases (Lieber, 2009b: 363) as in some parasyntetic compounds. Only subordinate compounds can have deverbal constituents as heads where the verb selects the non-head semantically as argument or as complement/adjunct (resulting in synthetic compounds). Lastly, it seems thus far as if only subordinate compounds can have semi-words as constituents, resulting in (neo-)classical compounds.

The third taxonomic rank pertains to headedness, defined on the morphosemantic level. Without being ignorant about the ongoing debate on headedness in morphology circles, we simply maintain Scalise and Bisetto’s (2009) definition and interpretation of the head as the semantic head of the compound (see also Booij, 1992). Whereas they indicate that all three major compound types can be both endocentric or exocentric (universally speaking), we claim that all compound types in Afrikaans and Dutch can be endocentric, but only the following can be exocentric:⁴

- (5) Subordinate, ground: $[[a]_V [b]_N]_V$ (Du. *knip+oog* **snip+eye** ‘to wink’), or $[[a]_V [b]_N]_N$ (Afr. *suip+lap* **booze+cloth** ‘drunkard’)
- (6) Attributive, ground: $[[a]_{Adj} [b]_N]_N$ (Afr. *rooi+kop* **red+head** ‘ginger (derogatory)’), or $[[a]_N [b]_N]_N$ (Du. *spleet+oog* **slit-eye** ‘Asian person (derogatory)’)⁵

Finally, another important addition to our taxonomy is the grouping of language specific/marginal cases (on the right-hand side of Figure 3). This choice should be understood in terms of the computational needs of this project, where one often needs a category for instances that do not fit the other main categories well. Such a category is currently called “Other” in CompoNet, and is used as a “dustbin” for anything that cannot be categorised as “Subordinate”, “Attributive”, “Appositive” or “Coordinate”. However, instead of having a very vague “Other” category, we try to be precise and explicit about these language specific categories. With regard to Afrikaans and Dutch, we identify three categories, viz. compounding compounds (Afr. *samestellende samestellings*), parasyntetic compounds (Afr. *samestellende afleidings*), and separable complex verbs (Afr. *samekoppelings*).

Compounding compounds are compounds that are formed with a noun as head, and either a NP (Adj+N, or Num+N) or PP (P+N) as non-head. Note that this is a specific kind of construction, and should as such not be confused with recursiveness in compounding. Unlike in subordinate, attributive and appositive phrasal compounds, the NP or PP in compounding compounds can only have two constituents. Also, a binary, left-branching interpretation of the compound as a recursive compound is impossible. Compare for instance a jocular example like Du. *gescheurde+broek+hersteller* **ripped+pants+repairer** ‘repairer of ripped pants’. If we would assume that *hersteller* first combined with *broek* to form *broekhersteller*, then a *gescheurde broekhersteller* would have been a ‘pants repairer who was ripped’. In other words, in compounding compounds, the compound as a whole is formed by means of the usual process of compounding (Booij and Van Santen, 1998: 179). In contrast,

⁴ Note that an example like Afr. *wag-n-bietjie* **wait-a-bit** ‘Buffalo Thorn (tree type)’ should not be analysed as exocentric, since it is actually a back-formation of *wag-n-bietjie-boom* **wait-a-bit-tree** ‘Buffalo Thorn’. There is a handful of highly lexicalised phrases (written concatenatively with hyphens, indicating their word status) that are exocentric, e.g. Afr. *een-twee-drie* **one-two-three** ‘quickly’, or Du. *vergeet-me-niet-je* **forget-me-not-DIM** ‘idem’. Most of these cases are names of plants, birds, food, etc., and in our opinion, are not productive in Afrikaans and Dutch. However, this will need to be established through future research.

Other problematic examples include highly lexicalised (metaphoric) compounds (like Du. *pad(den)+stoel* **frog+chair** ‘mushroom’), or simplexes that were diachronically speaking endocentric compounds (like Afr. *hard+loop* **fast+walk** ‘run’, which is still today considered an endocentric attributive compound in Dutch). For our purposes we consider both these cases as simplexes, but it could also be a theme for future research.

⁵ The compound *spleetoog* can also refer to a squinted eye, in which case it is endocentric.

in parasynthetic compounds, the compound is formed by means of derivation; compare for instance Du. *vijf+jaar-s* **five+year-extAdv** ‘five-yearly’, or Afr. *besluit+ne(e)m-ing* **decision+take-extN** ‘decision making’.

Lastly, we also include separable complex verbs in our taxonomy as a language specific category, and specifically as endocentric ground compounds. There is a vast literature on whether examples like Du. *op+zoeken* **up+look** ‘look up/search for’, and Afr. *af+sny* **off+cut** ‘cut off’ should be seen as compounds or not. Suffice to point the interested reader to Booij’s (2010) recent summary and discussion of the topic, and to state that we consider separable complex verbs as language specific compounds, based on the fact that they follow the same stress pattern as other compounds in Afrikaans and Dutch (i.e. main stress on the left-hand constituent).

4 Conclusion

In this paper, we have evaluated Bisetto and Scalise’s (2005) and Scalise and Bisetto’s (2009) compound taxonomies for purposes of revising the Dutch part of CompoNet, and to extend CompoNet by adding Afrikaans as a new language. Similar to Vercellotti and Mortensen’s (2012) critique of these taxonomies, we also suggested some changes, which might actually be more of interest to linguists. However, in our case we had a very practical aim as well, namely to explicate various aspects of the framework for practical analysis and annotation of Afrikaans and Dutch data in the CompoNet database. As is illustrated by Table 2 (in the appendix), we were able to comprehensively formalise the various patterns of compounding in Afrikaans and Dutch, and in the next phase of the project we will revise the original Afrikaans and Dutch data based on our taxonomy, in order to develop two supplementary databases (not part of the official CompoNet, but still using all their fields and conventions). Such databases could in future be used for comparative research, not only between Afrikaans and Dutch, but also with other languages in the CompoNet database.

Specific topics that need to be investigated in future include phrasal compounds (e.g. if we perhaps missed some patterns, what kind of phrases occur in which kinds of compounds?), exocentric compounds (i.e. do they only occur as ground compounds, or was our data skewed?; do we really need to include exocentricity in a compounding taxonomy for Afrikaans and Dutch), and (neo-)classical compounds (i.e. are (neo-)classical compounds always subordinate compounds?). Another topic pertains to the productivity of verbal compounds. Booij (2007: 92) states that Germanic languages do not have processes of verbal compounding, but that in Frisian occasionally new NV compounds do occur; we suspect that the same might be true for Afrikaans. Other topics of comparative research on compounding in Afrikaans and Dutch include whether (and why) Afrikaans has more A+N compounds than Dutch, the controversial topic of left-headed constructions in Dutch (e.g. Du. *kabinet-Zuma* **cabinet-Zuma** vs. Afr. *Zuma-kabinet* **Zuma-cabinet** ‘cabinet of (president) Zuma’), the difference of spreading of interfixes (linking morphemes), and differing stress patterns in compounding compounds in these two languages.

Acknowledgements

The Automatic Compound Processing (AuCoPro) project (<http://tinyurl.com/aucoopro>) was funded through a research grant from the Nederlandse Taalunie (Dutch Language Union) and the South African Department of Arts and Culture (DAC), as well as grants from the South African National Research Foundation (NRF) (grant number 81794), and the European Network on Word Structure (NetWordS) (European Science Foundation) (Grant number: 5570). Views expressed in this publication cannot be ascribed to any of these funding organisations. We would also like to acknowledge the contributions of Benito Trollip, who populated the first version of the Afrikaans section in the CompoNet database. Thank you also to the anonymous reviewers for their comments and suggestions.

References

- Giorgio F. Arcodia, Nicola Grandi and Fabio Montermini. 2009. Hierarchical NN compounds in a cross-linguistic perspective. *Rivista di Linguistica*, 21(1):11-33.
- Pierre J.L. Arnaud and Vincent Renner. 2014. English and French [NN]_N lexical units: A categorial, morphological and semantic comparison. *Word Structure*, 7(1):1-28.

- Antonietta Bisetto and Sergio Scalise. 2005. The classification of compounds. *Lingue e Linguaggio*, 4(2):319-332.
- Geert Booij. 1992. Compounding in Dutch. *Rivista di Linguistica*, 4:37-59.
- Geert Booij. 2007. *The grammar of words*. Oxford University Press, Oxford.
- Geert Booij. 2010. *Construction Morphology*. Oxford University Press, Oxford.
- Geert Booij and Ariane Van Santen. 1998. *Morfologie*. Amsterdam University Press, Amsterdam.
- Antonio Fábregas and Sergio Scalise. 2012. *Morphology: From Data to Theories*. Oxford University Press.
- Wim De Haas and Mieke Trommelen. 1993. *Morfologisch Handboek van het Nederlands [Morphologic Handbook of Dutch]*. SDU Uitgeverij.
- Walter Haeseryn, Kirsten Romijn, Guido Geerts, Jaap de Rooij & Maarten Cornelis van den Toorn. 1997. *Algemene Nederlandse Spraakkunst [ANS]*. Martinus Nijhoff, Groningen.
- Willem Kempen, 1969. *Samestelling, Afleiding en Woordsoortelike Meerfunksionaliteit in Afrikaans [Compounding, Derivation and Conversion in Afrikaans]*. Nasou, Cape Town, South Africa.
- Ronald Langacker. 2008. *Cognitive Grammar: A Basic Introduction*. Oxford University Press, New York.
- Rochelle Lieber. 2009a. A Lexical Semantic Approach to Compounding. In Rochelle Lieber and Pavol Štekauer, editors, *The Oxford Handbook of Compounding*, pages 78-104. Oxford University Press, Oxford.
- Rochelle Lieber. 2009b. IE, Germanic: English. In Rochelle Lieber and Pavol Štekauer, editors, *The Oxford Handbook of Compounding*, pages 34–53. Oxford University Press, Oxford.
- Diarmuid Ó Séaghdha. 2008. *Learning compound noun semantics*. Ph.D. thesis, University of Cambridge, Cambridge, UK.
- Sergio Scalise and Antonietta Bisetto. 2009. The classification of compounds. In Rochelle Lieber and Pavol Štekauer, editors, *The Oxford Handbook of Compounding*, pages 34–53. Oxford University Press, Oxford.
- Gerhard Van Huyssteen. 2004. Motivating the composition of Afrikaans reduplications: a cognitive grammar analysis. In Günter Radden and Klaus-Uwe Panther, editors, *Studies in Linguistic Motivation*, pages 269-292. Mouton de Gruyter, Berlin.
- Gerhard Van Huyssteen. 2014. Morfologie [Morphology]. In Wannie Carstens and Nerina Bosman, editors, *Kontemporêre Afrikaanse Taalkunde [Contemporary Afrikaans Linguistics]*, pages 171-208. Van Schaik Uitgewers, Pretoria, South Africa.
- Mary Lou Vercellotti and David R. Mortensen. 2012. A classification of compounds in American Sign Language: an evaluation of the Bisetto and Scalise framework. *Morphology*, 22(4):545-579.
- Ben Verhoeven, Menno van Zaanen, Walter Daelemans and Gerhard Van Huyssteen. 2014. Automatic Compound Processing: Compound Splitting and Semantic Analysis for Afrikaans and Dutch. In *Proceedings of the First Workshop on Computational Approaches to Compound Analysis (ComAComA)*, Dublin, Ireland.

Appendix

| | | | | | |
|-------------|---------|---|----------------------------------|---------------------------------|-----------------------|
| Subordinate | VerbNex | [[a] _N [[b] _V extAdj] _{Adj} Adj _i | <i>hand+vervaardig-de</i> | hand+produce-extAdj | hand-made |
| | | [[a] _N [[b] _V extN] _N N | <i>gras+sny-er</i> | grass+cut-extN | lawn mower |
| | | [[a] _V [[b] _V extN] _N N | <i>eet+sta(a)k-ing</i> | eat+strike-extN | hunger strike |
| | | [[a] _{Adj} [[b] _V extN] _N N | <i>kaal+nael-er</i> | naked+run-extN | streaker |
| | G | [[a] _{Adv} [[b] _V extAdj] _{Adj} Adj _i | <i>dig+bebos-te</i> | thick+afforest-extAdj | thickly wooded |
| | | [[a] _N [b] _N N | <i>tafel+poot</i> | table+leg | table leg |
| | | [[a] _N [b] _{Adj} Adj _i | <i>kleur+blind</i> | colour+blind | colour blind |
| | | [[a] _P [b] _N N | <i>buite+kamer</i> | outside+room | outside room |
| | | [[a] _{Num} [b] _N N | <i>twee+klank</i> | two+sound | diphthong |
| | | [[a] _V [b] _N N | <i>stryk+plank</i> | iron+board | ironing board |
| | | [[a] _N [b] _V V | <i>raad+pleeg</i> | advice+commit | consult |
| | Ph | [[a] _{VP} [b] _N N | <i>skop-skiet-en-donder-film</i> | kick-shoot-and-hit-movie | action movie |
| | | [[a] _{NP} [b] _N N | <i>kaas-en-wyn-onthaal</i> | cheese-and-wine-party | cheese and wine party |
| | NeoC | [[a] _{Sw} [b] _{Sw} N | <i>hidro+logie</i> | hydro+logy | hydrology |
| | | [[a] _{Sw} [b] _N N | <i>bio+brandstof</i> | bio+fuel | biofuel |
| | | [[a] _N [b] _{Sw} N | <i>Japan(n)+(o)logie</i> | Japan+ology | Japanese studies |
| Attributive | G | [[a] _{Adj} [b] _N N | <i>blou+draad</i> | blue+wire | galvanised wire |
| | | [[a] _{Num} [b] _N N | <i>tien+kamp</i> | ten+camp | decathlon |
| | | [[a] _{Adv} [b] _N N | <i>terug+weg</i> | back+way | the way back |
| | | [[a] _{Adv} [b] _{Adj} Adj _i | <i>donker+blond</i> | dark+blonde | dark blonde |
| | | [[a] _{Num} [b] _{Adj} Adj _i | <i>twee+maandeliks</i> | two+monthly | bimonthly |
| | Ph | [[a] _{AP} [b] _N N | <i>los-en-vas-praatjies</i> | loose-and-set-talks | random chatting |
| | | [[a] _{NP} [b] _N N | <i>kop-aan-kop-botsing</i> | head-on-head-collision | head-on collision |
| | | [[a] _{PP} [b] _N N | <i>in-die-lug-vraag</i> | in-the-air-question | rhetorical question |
| | G | [[a] _N [b] _N N | <i>treffer+liedjie</i> | hit+song | hit song |
| | | [[a] _N [b] _{Adj} Adj _i | <i>yster+sterk</i> | iron+strong | strong as iron |
| | | [[a] _V [b] _{Adj} Adj _i | <i>spring+lewendig</i> | jump+lively | alive and well |
| | | [[a] _P [b] _{Adj} Adj _i | <i>deur+nat</i> | through+wet | soaked |
| Appositive | Ph | [[a] _{VP} [b] _{Adj} Adj _i | <i>kielie-my-maag-lekker</i> | tickle-my-stomach-nice | idem |
| | | [[a] _{NP} [b] _{Adj} Adj _i | <i>sonsak-in-Ibiza-mooi</i> | sunset-in-Ibiza-pretty | idem |
| Coord | G | [[a] _N [b] _N N | <i>skrywer-boer</i> | writer-farmer | writer-farmer |
| | | [[a] _{Adj} [b] _{Adj} Adj _i | <i>stom+verbaas</i> | mute+surprised | very surprised |
| | | [[a] _V [b] _V V | <i>sit+lê</i> | sit-lie | sit and lie |
| | | [[a] _P [b] _P P | <i>voor+op</i> | before+above | first |
| CC | Ph | [[[a] _{Adj} [b] _N NP[c] _N N] | <i>sosiale+sekerheid(s)+reg</i> | social+security+law | social security law |
| | | [[[a] _{Num} [b] _N NP[c] _N N] | <i>twee+sitplek+motor</i> | two+seat+car | two-seater |
| | | [[[a] _P [b] _N PP[c] _N N] | <i>buite+boord+motor</i> | out+board+motor | outboard motor |
| SCV | G | [[a] _P [b] _V V | <i>in+gooi</i> | in+throw | throw in |
| | | [[a] _{Adv} [b] _V V | <i>neer+gooi</i> | down+throw | throw down |
| | | [[a] _N [b] _V V | <i>vleis+braai</i> | meat+roast | barbeque |
| Para Synth | Ph | [[a] _{PP} extN] _N | <i>ter+aarde+bestel(l)-ing</i> | to+earth+deliver-extN | burial |
| | | [[a] _{NP} extN] _N | <i>groot+skaal-s</i> | large+scale-extAdj | large-scale |
| | | [[a] _{VP} extN] _N | <i>alleen+lo(o)p-er</i> | alone+walk-extN | loner |
| | G | [[a] _{Adj} [b] _N extAdj] _{Adj} | <i>blou+kleur-ig</i> | blue+colour-extAdj | blue-coloured |
| | | [[a] _{Num} [b] _N extAdj] _{Adj} | <i>een+bla(a)r-ig</i> | one+leaf-extAdj | monopetalous |

Table 2. Prototypical endocentric compounds in Afrikaans and Dutch (with Afrikaans examples)⁶

⁶ Abbreviations: VerbNex=verbal-nexus; G=ground; Ph=phrasal; NeoC=(neo-)classical; Coord=coordinate; CC=compounding compound; SCV=separable complex verb; ParaSynt=parasynthetic