Constructions and Lexical Units: An Analysis of Dutch Numerals

Geert Booij

1 Construction morphology

The theory of Construction Morphology (*CM*) which is the theoretical framework of this article (Booij 2005, Booij 2007, Booij 2009) aims at a proper understanding of the relation between syntax, morphology, and the lexicon, and at providing a framework in which both the differences and the commonalities of word level constructs and phrase level constructs can be accounted for.

There are two basic approaches to the linguistic analysis of complex words. In the morpheme-based approach which was dominant in post-Bloomfieldian American linguistics, a complex word is seen as a concatenation of morphemes. In this approach, morphological analysis can be defined as the 'syntax of morphemes'. For instance, the English word *walker* can be seen as a concatenation of the verbal morpheme *walk* and the nominalizing suffix *-er* that carries the meaning 'agent'. Alternatively, we might take a word-based perspective in which words are the starting points of morphological analysis which is done by comparing sets of words like:

(1)

buy	buyer
eat	eater
shout	shouter
walk	walker

and conclude to a formal difference between the words in the left and the right column that correlates systematically with a meaning difference: the words on the right in (1) have an additional sequence *-er* compared to those on the left, and denote the agent of the action expressed by the verbs on the left. This paradigmatic relationship between *buy* and *buyer* can be projected onto the word *buyer* in the form of word-internal morphological structure:

(2) $[[\text{buy}]_V \text{ er}]_N$

Moreover, the set of words listed in (1) may give rise, in the mind of the speaker of English, to an abstract schema of the following form:

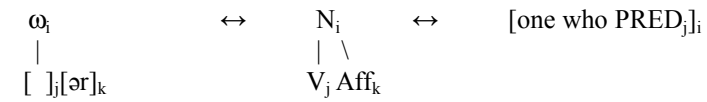
(3) $[[\text{x}]_V \text{ er}]_N$ 'one who Vs'

This schema expresses a generalization about the form and meaning of existing deverbal nouns in *-er* listed in the lexicon, and can also function as the starting point for coining new English nouns in *-er* from verbs. That is, new deverbal nouns in *-er* are not necessarily coined on analogy with a specific existing

deverbal word in *-er*, but may be formed on the basis of this abstract schema. A new word is formed by replacing the variable *x* in the schema with a concrete verb. This is the operation of 'unification'. For instance, the recently coined English verb *to skype* 'to communicate by means of Skype' can be unified with schema (3), resulting in the new noun *skyper*.

The form part of schema (3) is a combination of morphological form and phonological form. The morphological form is $[\text{V Affix}]_{\text{N}}$, and the phonological form is that of a prosodic word that ends in the sound sequence /ər/. The relation between the three levels of representation (phonological form, morpho-syntactic form, and semantics) can be represented more explicitly as follows, assuming the tripartite parallel architecture of the grammar proposed in (Jackendoff 2002):

Figure 1. The schema for deverbal *-er*.



The symbol ↔ denotes the correlations between the three levels of representation; corresponding properties of subparts are expressed by co-indexation. The computation of the actual phonological form of individual complex words is achieved by an interface module, that relates morphological information and phonological form. In the case of a word like *skyper* the computation is rather simple. In order to derive the prosodic word $(\text{skaj.pər})_{\omega}$ (the dot indicates a syllable boundary) on the basis of the phonological form of *skype* and that of the affix, the only 'interface' information we need is that the suffix *-er* is a 'cohering suffix' that fuses prosodically with its stem into one prosodic word. Thus, for the computation of the prosodic structure of *skyper*, we must ignore the morphological boundary. Hence, the word-internal syllable boundary does not coincide with the word-internal morphological boundary. This contrasts with the behaviour of non-cohering affixes such as the English prefix *un-*. In a word like *unable* the word-internal morphological boundary coincides with a syllable boundary: *un.a.ble* instead of *u.na.ble* (Booij & Rubach 1984).

It will be clear that in many cases the computation of the phonological form of a complex word is more complicated. Moreover, affixation is not the only form of constructing complex words. Think, for instance, of reduplication, or of subtractive morphology where morphological information is expressed by deleting certain segments. However, in this article, I will focus on the analysis of the non-phonological properties of lexical constructs.

The idea that word formation patterns can be seen as abstractions across sets of related words is rooted in a venerable tradition. For instance, the German linguist and Junggrammatiker Hermann Paul wrote in his famous *Prinzipien der Sprachgeschichte*, published in 1880, that the language learner will start with learning individual words and word forms, but gradually (s)he will abstract away from the concrete words (s)he has learned, and coin new words and word forms according to abstract schemas. This enables the language user to be crea-

tive in word formation and inflection (Paul 1880 [3rd edition 1898]), p. 102). This tradition is continued in the paradigmatic approach to word formation in Europe (Schultink 1962, van Marle 1985), and in recent work in varieties of non-transformational generative grammar such as Head-driven Phrase Structure Grammar (Riehemann 1998, Riehemann 2001).

Since such schemas depend on relationships between words, this morphological model has been called the network model (Bybee 1995), and the notion 'network' is indeed a proper term for conceptualizing the set of relationships between words in a lexicon (Bochner 1993). This approach may also be qualified as the 'abstractive' approach (Blevins 2006) because the coinage of new words depends on abstractions over the sets of existing words and word forms in the lexicon of a language.

Schema (3) may be said to license the individual deverbal nouns in *-er* in the English lexicon. Complex words, once they have been coined will be stored in the lexicon of a language (which generalizes over the lexical memories of the individual speakers of that language) if they have idiosyncratic properties and/or they have become conventionalized.

CM assumes that complex words, i.e. the outputs of morphological operations, can be listed in the lexicon. Morphological schemas therefore have two functions: they express predictable properties of existing complex words, and indicate how new ones can be coined (Jackendoff 1975). This conception of the grammar avoids the well known rule/list fallacy (Langacker 1987), the unwarranted assumption that linguistic constructs are either generated by rule or listed, and that being listed excludes a linguistic construct from being linked to a rule at the same time.

The relation between schema (3) and the individual words that conform to this schema is that of 'instantiation': each of the nouns in *-er* listed in (1) instantiates the schema in (3). Schema (3) provides a direct account of the fact that *-er* is a bound morpheme that does not occur as a word by itself, since this morpheme is not listed in the lexicon. Its existence is bound to its occurrence in schema (3). The same sequence of sounds /ər/ is used in other morphological schemas as well, for instance in the schema for the comparative form of English adjectives.

The notion construction (defined as a pairing of form and meaning) is a traditional notion used in thousands of linguistic articles and books. In most cases it refers to a syntactic pattern in which particular formal properties correlate with specific semantics. For instance, many linguists of English speak of 'the passive construction' since sentences with passive meaning in English have a specific syntactic form.

A well known example of a syntactic construction is the caused motion construction exemplified by sentence (4) (Goldberg 2006: 73):

(4) Pat sneezed the foam off the cappucino

In this sentence, the verb *to sneeze* is used as a transitive verb, although it is normally an intransitive verb. Its use as a transitive verb correlates with the presence of an object that moves along a path specified by a PP. The meaning component that the sneezing caused the foam to move is therefore to be seen as a property of the caused motion construction as a whole.

The notion 'construction' plays an important role in a number of recent linguistic models: Construction Grammar (Croft 2001, Fried & Östman 2004, Goldberg 1995, Goldberg 2006), the Simpler Syntax Model (Culicover & Jackendoff 2005, Culicover & Jackendoff 2006), Cognitive Linguistics (Langacker 1999), and Head-driven Phrase Structure Grammar (HPSG) (Sag 2007, Sag, Wasow & Bender 2003). The following features of the constructional approach are of high relevance for the further articulation of *CM*:

(5) "Pieces of syntactic structure can be listed in the lexicon with associated meanings, just as individual words are; these are the MEANINGFUL CONSTRUCTIONS of the language."
"Construction grammar makes no principled distinction between words and rules: a lexical entry is more word-like to the extent that it is fully specified and more rule-like to the extent that it contains variables [...]."
"Lexical entries are arranged in an inheritance hierarchy." (Jackendoff 2008: 15).

Constructions can vary in size and complexity, as illustrated in the following sketch of the syntax-lexicon continuum by Croft:

Table 1. The syntax-lexicon continuum (Croft 2001: 17)

Construction type	*Traditional name*	*Examples*
Complex and (mostly) schematic	syntax	[SBJ be-TNS VERB-en by OBL]
Complex and (mostly) specific	idiom	[pull-TNS NP's leg]
Complex but bound	morphology	[NOUN-s], VERB-TNS]
Atomic and schematic	syntactic category	[DEM], [ADJ]
Atomic and specific	word / lexicon	[this], [green]

In this article I argue that regularities in the structure and formation of both complex words and phrasal lexical units can be insightfully accounted for by making use of the notion construction. I illustrate this by giving an analysis of the construction of Dutch numerals.

2 Dutch numerals

Consider the following numerals of Dutch and their glosses in English:

(6) a. 5 vijf 'five'
b. 15 vijf-tien 'five-teen'

c. 51 een-en-vijftig ‘one-and-fifty’
d. 105 honderd (en) vijf ‘one hundred (and) five’

Most numerals of Dutch and English are complex linguistic expressions, formed by a recursive system of rules that enables the language user to form an in principle infinite set of numerals. In Dutch, as in English and German, all numerals above the number 12 are such complex expressions. The numeral *vijf-tien* ‘15’ in (6b) has the shape of a compound consisting of two lexeme constituents, *vijf* ‘5’ and *tien* ‘10’. The next two examples (6c–d), on the other hand, have the appearance of phrases, formed by means of coordination with the conjunction *en* ‘and’. The difference between (6c) and (6d) is that only in (6d) can the conjunction be omitted. Another difference between (6c) and (6d) is that in (6c) the conjunction *en* is pronounced as [ən], whereas in (6d) it must be pronounced as [ɛn].

Even though (6c) and (6d) look like phrases, they can function as bases of word formation, for the formation of ordinal numerals by means of the suffixes *-ste* and *-de*:

(7) a. een-en-vijftig-ste ‘one-and-fifty-th, fifty-first’
b. honderd(-en)-vijfde ‘hundred (and) fifth’

One may therefore conclude either that the numerals (6c–d) are words, or that morphological operations can take phrases as their bases. This issue is taken up in section 4.

Thus, Dutch numerals raise the question to what extent their construction is a matter of morphology, and to what extent it belongs to syntax. The reference grammar of Dutch, the *Algemene Nederlandse Spraakkunst* (Haeseryn, Romyn, Geerts, de Rooij & van den Toorn 1997) discusses the formation of numerals under the heading of word classes and word formation, whereas Hurford, in his studies of numerals (Hurford 1975, Hurford 1987, Hurford 2003, Hurford 2007) considers them as syntactic constructs, most explicitly in Hurford (2007: 777). In this paper I argue that Dutch numerals are a mixed bag of derived words, compounds, and syntactic idioms, and form a complex network of constructional schemas of varying degrees of abstractness.

Section 3 deals with the numerals that are usually referred to as cardinal numerals. Note, however, that such words that denote numbers have more functions than expressing the cardinality of sets. Wiese (2007: 759–60) distinguishes three functions:

(a) cardinal number assignments, as in *three books*; *I saw three of the students*;
(b) ordinal number assignment, as in *group 5*, (*the year*) *2001*, *Downing Street 10*;
(c) nominal number assignment, as in *line 5* (name of one of a number of lines in public transport), *MasterCard # 6666 etc*. (number of credit card).
(d) In addition, cardinal numerals can be used for counting: *one, two, three ….*

When used as cardinal numbers, these numerals function attributively, as in *three books*, or as heads of phrases, as in *I saw three of the students*. In the uses (b) and (c) the numeral follows the head noun of a phrase that functions as a proper name.

In section 4 I deal with the morphological construction of ordinal numerals and in section 5 with the construction of fraction names. Section 6 summarizes my findings and theoretical conclusions.

3 Cardinal numerals

Let me first give a representative selection of the expressions that are used as cardinal numerals in Dutch:

(8) a. simplex numerals 1–12

1 een [e:n]
2 twee
3 drie
4 vier
5 vijf
6 zes
7 zeven
8 acht
9 negen
10 tien
11 elf
12 twaalf;

b. numerals 13–19

13 der-tien
14 veer-tien
15 vijf-tien
16 zes-tien
17 zeven-tien
18 acht-tien
19 negen-tien;

c. numerals 20–90

20 twin-tig
30 der-tig
40 veer-tig
0 vijf-tig
0 zes-tig
70 zeven-tig
80 tach-tig
90 negen-tig

d. numerals 21–99

21 een-en-twin-tig
32 twee-en-der-tig
43 drie-en-veer-tig
54 vier-en-vijf-tig
65 vijf-en-zes-tig
76 zes-en-zeven-tig
87 zeven-en-tach-tig
99 negen-en-negen-tig

e. numerals 100 and higher

100	honderd
101	honderd (en) een
1,000	duizend
1,000,000	miljoen
1,000,000,000	miljard

The numeral *een* [e:n] has the same orthographic form as the indefinite singular determiner *een* [ən]. Historically, the determiner derives from the numeral. A characteristic difference between grammatical words and words of lexical categories is that only the former can have schwa [ə] as their only vowel (Booij 1995). The difference in phonetic form between the two lexemes *een* reflects this phonological constraint.

There is some allomorphy to be observed in (8b) in comparison with the forms in (8a): *drie* has the allomorph *der-*, and *vier* has the allomorph *veer-*. The suffix for the numerals 20–90, the multiples of 10, is *-tig /təɣ/*; it is affixed to the allomorphs *twin-* for *twee*, *der-* for *drie*, *veer* for *vier*, and *tach-* for *acht*, as shown in (8c). The suffix *-tig* has also been reanalysed as a word, as in *tig keer* 'for the umpteenth time'. In that case, the vowel is realized as [ɪ] because a word of lexical category cannot have schwa [ə] as its only vowel.

The numerals listed in (8c) are clear cases of complex words. The following morphological schema express the commonalities of this closed set of words:

(9) $[[x]^{i}_{\text{Num, [dig]}}\ \text{tig}]^{j}_{\text{Num}} \leftrightarrow [\text{NUM}^{i} \times 10]^{j}$

(The feature [dig] denotes the set of digital numbers 1–9). Schema (9) represents the correlation between a particular form and a particular meaning. the meaning of each numeral in *-tig* is 10 times the value of the digit number. The superscripts *i* and *j* identify the form and the corresponding meaning (= arithmetical value) of the base and the complex numeral. The indices *i* and *j* are variables for lexical indexes: each lexical unit in the lexicon carries its own lexical index. NUM stands for the arithmetical value of a numeral. Even though this schema is unproductive, we need it because we want to express the predictable properties of this set of words.

The individual words in *-tig* are morphological constructs that instantiate this schema. They can be formed by unification of the base digit numerals with the schema. They inherit the properties specified by the schema unless they are specified otherwise. This is the idea of default inheritance discussed in detail in (Briscoe, Copestake & de Paiva 1993, Evans & Gazdar 1996, Kilbury, Petersen & Rumpf 2006, Lascarides & Copestake 1999): the specification of a word for a particular property is inherited from the dominating node in the hierarchical lexicon, unless the actual lexical entry has another specification for that property. The mechanism of default inheritance is necessary because we want to be able to express that a word has an exceptional property, although it is regular in most respects.

It is obvious that all numerals in *-tig* are stored in the lexicon. Hence, the schema is to be seen as a redundancy statement that specifies which information concerning the individual numerals in *-tig* is predictable (Jackendoff 1975). The words *twin-tig, der-tig, veer-tig*, and *tach-tig* are exceptional since their base is not an existing numeral. So the information concerning the base part of these numerals does not count as predictable information, as it does not carry a lexical index of its own.

We might represent these numerals as follows in a hierarchical lexicon, linked to the schema (9) (with (arbitrary) lexical indices such as 82).

(10) $[\text{vijf}]^{82}{}_{\text{Num, [dig]}}\ \text{tig}]^{83}{}_{\text{Num}} \leftrightarrow [5^{82} \text{ x } 10]^{83}$
$[\text{veer}]_{\text{Num, [dig]}}\ \text{tig}]^{84}{}_{\text{Num}} \leftrightarrow [4 \text{ x } 10]^{84}$

The lexical index indicates that the form and meaning of *vijf* recur in *vijftig*. Thus, co-indexation expresses the part-of-relation that exists between a complex word and its base. The base *veer* is not a numeral of Dutch, and hence it does not carry a lexical index that identifies it with an independently existing lexeme in the lexicon.

The arithmetical operations involved in the formation of complex numerals of Dutch are addition and multiplication. The default linguistic expression of addition is coordination, with optional use of the conjunction *en* 'and'. Traces of this universal syntactic mechanism can be seen in the formation of numerals, but in Dutch numerals it is grammaticalized into three specific patterns listed and exemplified in (11); these patterns impose specific restrictions on the presence of an overt conjunction and the order of the subconstituents:

(11) a. no overt conjunction, lower number before higher number: *vijf-tien* '15';
b. conjunction, lower number before higher number *een-en-vijf-tig* '51';
c. optional conjunction before the last numeral, higher number before lower number: *honderd (en) vijf* '105', *twee-duizend acht-honderd (en) vijf* '2805'.

These data suggest that these patterns are specific constructions that reflect the general syntactic principles of coordination of Dutch, but need to be stated separately as specific instantiations of the coordinating construction, with properties of their own. That is, they are idioms. These numerals may contrast with the regular pattern of coordination, as illustrated by the following minimal pair:

(12) a. vijf-tien boeken 'fifteen books'
b. vijf en tien boeken 'five and ten books'

In (12a) the phrase denotes one set of books with cardinality 15, whereas (12b) denotes two different sets of books with the cardinalities 5 and 10 respectively. A numeral like *vijftien* has the appearance of a compound word. It has the stress pattern of Dutch compounds, with main stress on the first constituent. Yet, it does not possess the properties of regular compounds of Dutch such as being right-headed. In *vijftien* the right constituent *tien* 'ten' has no semantic head properties with respect to the word as a whole. This is explained by considering this special type of compound word as being derived historically from (asyndetic) coordination. It sides with the regular right-headed compounds of Dutch, however, in that main stress is on the first constituent: *víjf-tìen*.

Dutch (and closely related Germanic languages such as Frisian and German) differ from English as to the order of the number constituents below 100. In

English the higher number comes before the lower number after 20 (*twenty-one*, etc.), whereas in Dutch this switch of order takes place after 100. The exact locus of switch varies from language to language. In Italian, for instance, the switch takes place after 16: *se-dici* '16', but *dici-a-sette* '17'.

These languages all conform to the following universals proposed by (Greenberg 1978: 273):

(13) "26. If in a language, in any sum the smaller addend precedes the larger, then the same order holds for all smaller numbers expressed by addition.
27. If in a language, in any sum the larger addend precedes the smaller, then the same order holds for all larger numbers expressed by addition."

The second arithmetical operation involved in the construction of Dutch numerals is that of multiplication, as in the following numerals:

(14)
a.	vijf-tig	5 x 10
b.	vijf-honderd	5 x 100
	vijf-duizend	5 x 1000
c.	vijf-miljoen	5 x 1,000,000
	vijf-miljard	5 x 1,000,000,000

The words of the type *vijftig* are discussed above. The words *honderd* 'hundred', *duizend* 'thousand', *miljoen* 'million' and *miljard* 'billion' can be considered a subset of the category of measure nouns. Hurford refers to these units that function as the basis of multiplication with the symbol M (Hurford 1975, 1987). They have properties of nouns, as shown by the fact that they can be pluralized, as in:

(15)

a. honderd-en bezoeker-s
hundred-PL visitor-PL
'hundreds of visitors'

b. duizend-en gulden-s
thousand-PL guilder-PL
'thousands of guilders'

c. miljoen-en sterr-en
million-PL star-PL
'millions of stars'

d. miljard-en boek-en
billion-PL book-PL
'billions of books'

In numerals these multiplication bases have the singular form, just like other Dutch measure nouns such as *meter* 'metre' and *kilo* 'kilogram':

(16)

a. drie meter
three meter.SG
'three meters'

b. vijf kilo
five kilo.SG
'five kilograms'

The numerals *honderd* en *duizend* are noun-like in that they can be pluralized. However, they differ from *miljoen* and *miljard*, since they can be used without a preceding numeral, unlike *miljoen* and *miljard* (except in headlines of news papers, which have a special syntax):

(17) a. honderd / duizend boeken ‘hundred / thousand books’
b. *miljoen / een [e:n] miljoen boeken ‘one million books’
*miljard / een [e:n] miljard gulden ‘one billion guilders’

Thus, we need the following specifications for these classes of words:

(18) a. honderd, duizend: [+N, + Num, +M]
b. miljoen, miljard: [+N, +M]

The feature [+N] predicts that these words, being nouns, can be pluralized. The feature [+M] qualifies them as measure nouns that appear in their singular form after a numeral. The words in (18b) do not carry a lexical feature [+Numeral], and hence they cannot occur by themselves as numerals in noun phrases.

Numerals can project a phrase that I will denote by means of the label NumP. Numeral phrases are a subcategory of quantifier phrases, but I will refrain from a detailed discussion of the complexities of the structure of noun/determiner phrases and their quantifier subparts. Suffice it to mention here that numerals can project phrases since they can be modified:

(19) ongeveer twintig ‘approximately 20’
ruim twintig ‘amply 20’
om en nabij twintig ‘lit. around and close to 20, about 20’
meer dan twintig ‘more than 20’

The next issue is how to properly characterize the grammar of complex numerals. The productive schemas for Dutch are those for multiplication and addition. The multiplication schema is an instantiation of the general schema for constructions with a numeral followed by a measure noun that creates quantifying expressions. The additional property of this schema is that it specifies these expressions as numerals which can hence form part of larger complex numerals, as discussed below. The numeral *drie honderd* can be embedded in a numeral coordination structure, as in *driehonderd-en-vijf* ‘305’, unlike other measure expressions like *twee meter* (**twee meter en vijf*):

(20) *Multiplication schema*
$[Num^{i}\ Num^{j}_{[+M]}]^{k}_{Num} \leftrightarrow [NUM^{i} \times NUM^{j}]^{k}$
(where NUM stands for the arithmetical value of the corresponding formal constituents indexed as *i* and *j*).

Numerals like *honderd, duizend, miljoen,* and *miljard* are specified as belonging to the subcategory M, and hence they form bases of multiplication. This schema will generate numerals like the following:

(21) $[[\text{achttien}]_{Num}[\text{honderd}]_{Num}]_{Num}$ ‘eighteen hundred’
$[[\text{negen-en-negen-tig}]_{Num}[\text{honderd}]_{Num}]_{Num}$ ‘ninety-nine hundred’
$[[\text{honderd}]_{Num}[\text{duizend}]_{Num}]_{Num}$ ‘hundred thousand’
$[[[\text{twee}]_{Num}\ [\text{honderd}]_{Num}]_{Num}\ [\text{miljoen}]_{Num}]_{Num}$ ‘two hundred million’

As the last example illustrates, the schema can be applied recursively: the numeral *tweehonderd* contains an M (*honderd*) and modifies an M (*miljoen*).

An important constraint on the construction of these multiplication numerals is that the value of the modifying numeral must be lower than that of M. This constraint accounts for the difference in wellformedness between:

(22) a. [negen-en-negen-tig]-honderd / negen-duizend (en) negen-honderd '9900'
b. *[honderd en twee]-honderd / tien-duizend twee-honderd '10200'

The illformedness of the first option in (22b) reflects the universal global constraint on numeral formation referred to as the Packing Strategy principle. This principle, proposed in the work of Hurford on numerals, is meant to make a choice between different structural options that the language system provides (Hurford (2007: 774)). It basically says that you must first use the highest measure noun possible, so that you cover the largest subset possible with one numeral constituent. As (22a) with its two options show, the principle does not apply without exceptions. However, the numerals for multiples of 10 which end in *-tig* 'ten' cannot occur as modifiers of *hundred* unless they are preceded by a one digit numeral. So we get the following difference in expression possibilities for the numbers 9900 and 9000:

(23) 9900: negen en negentig-honderd *or* negen-duizend (en) negen-honderd
9000: *negentig-honderd / negen-duizend

Expressing the number 9000 as *negenduizend* is clearly the most economical option, and in line with the Packing Strategy. In the case of 9900 on the other hand, the two options do not differ substantially in terms of complexity although the first option violates the Packing Strategy constraint. Hence, economy of expression is another factor in choosing between different structural options. According to Hurford (2007: 779), the Packing Strategy may be explained as the effect of two pragmatic principles applied in counting entities: "Go as far as you can with the resources you have", and "Minimize entities".

As to the operation of addition for the formation of numerals, Dutch requires two schemas of coordination. The most general one is that for all numerals from 100 onwards:

(24) Addition schema for numeral expressions > 100
$[\text{Num}_C\text{*}\ ((\varepsilon n)\ \text{Num}_D)]^j_{\text{Num}} \leftrightarrow [\text{NUM}_C + \text{NUM}_D \ldots]^j$
(where NUM stands for the value of the corresponding Num.)

Num_C stands for numerals ≥ 100, and Num_D stands for numerals < 100. The asterisk indicates that this constituent is recursive, and can be repeated.

Schema (24) is an instantiation for numerals of the Dutch coordination construction in which constituents of the same category can be conjoined, and thus create a constituent of the same category. In the case of numerals, the conjunction is absent between numerals ≥ 100, and optional before Num_D, and in this

respect numerals differ from other conjoined categories. For instance, one cannot conjoin two Ns without an overt conjunction (compare *vader en moeder* 'father and mother' with **vader moeder* 'father mother').[1] This is why the subcase of numeral coordination requires a schema of its own. The semantics is that of addition, as specified in the schema. The schema will generate numerals like the following:

(25) [drie-miljoen] [vier-honderd-acht-duizend] [drie-honderd (en) twee]
'3,408,302'

If Num_D is ≥ 21, we might get two occurrences of the conjunction *en* (compare 26a and 26b). Although this does not lead to ungrammaticality, there is a preference for having *en* only before the last number constituent. Thus, in coining the numeral for the number 399, of the following two options, the first one is usually preferred:

(26) a. driehonderd negen-en-negentig '399'
b. driehonderd en negen-en-negentig '399'

The order of the numeral constituents under addition (coordination) as specified by schema (24) must be such that a coordinated numeral constituent with a higher value precedes a coordinated numeral constituent with a lower value within the addition scheme, in line with the Packing Strategy, as illustrated in (27).

(27) 5,002,600 *zes-honderd twee-duizend vijf-miljoen
vijf-miljoen twee-duizend zes-honderd
9900 *negenhonderd en negen duizend
negen-duizend (en) negenhonderd

Fixation of word order in coordinated structures as illustrated above is not an exclusive property of numerals. It is also found in binomial expressions like *salt and pepper* or *father and son*, and similar binomial expressions in many languages.

A second schema of addition is needed for the numerals 21–99, because there is a special order for such numerals in languages like Dutch and German, in which they differ from English. As mentioned above, these numerals are special in that the digit numeral for 1–9 precedes the numeral for the (multiple of) ten. In addition, the conjunction *en* is obligatorily present – unlike what is the case for numerals > 100 –, and it is not pronounced as [ɛn], but as [ən]. This syllable forms one prosodic word with the preceding numeral, and this prosodic structure (two adjacent vowels within the same prosodic word) triggers homorganic glide insertion (Booij 1995). Hence, we have Dutch numerals such as the following:

(28) 23: drie-en-twintig $(dri.jən)_ω(tʊɪn.təx)_ω$
62: twee-en-zestig $(tʊe:jən)_ω(sɛs.təx)_ω$

[1] Proper names can be conjoined, however, without overt conjunction, as in Jan-Peter 'John Peter' and Rijn-Schelde-kanaal 'Rijn-Schelde canal', which suggest that we need a specific subschema of proper name conjunction.

The relevant schema for the construction of these numerals is

(29) Schema for numerals between 21 and 99
$[Num^{i}_{Digit}$ ən $Num^{j}_{D}]^{k}_{Num} \leftrightarrow [NUM^{i} + NUM^{j}]^{k}$
(Num_{Digit} = 1–9, Num_{D} = 20, 30 .. 90).

The facts of Dutch numerals, as analysed above, thus confirm the general conclusion in Hurford (1987: 303) that "numeral constructions in all languages tend to be […] syntactic idioms". However, not all Dutch numerals are syntactic in nature synchronically. The numerals for 1–20 and for the multiples of 10 are clearly words. The numerals formed according to schema (29) might be considered words. The full vowel [ɛ] of the conjunction *en* has been reduced to schwa. Thus, this *en* has become a linking element. In this respect they are now like other types of compound: Dutch compounds may have [ən] as a linking element, as in *boek*-[ən]-*plank* 'book shelf'.

As to numerals > 100, they can be considered phrases, because they allow for the appearance of the conjunction *en* in its unreduced form, and hence have the form of syntactic coordination. The specific coordination pattern involved is lexicalized, however, in the sense that it applies to numerals only. There is a wealth of evidence that certain types of phrases must be specified in the lexicon (Booij 2002, Jackendoff 2002). The different schemata can be unified, and thus account for the construction of complex numeral expressions.

The formal restrictions on numerals mentioned above illustrate how the grammar of a language can restrict the use of general principles of syntactic construction for specific subdomains. Such grammaticalized conventions can also be observed in the use of numerals for specific purposes, such as naming years. Both in English and in Dutch the word *hundred/ honderd* can be omitted in names for years, under the condition that there is no overt conjunction:

(30) 1654: zestien-honderd (en) vier-en-vijf-tig / zestien (*en) vier-en-vijf-tig
sixteen hundred fifty-four / sixteen fifty-four

Moreover, we cannot use the alternative expression *duizend zeshonderd vier-en-vijf-tig* for denoting the year 1654. This shows how discourse-specific conventions may regulate and restrict the use of structural options offered by the language system (cf. Östman 2005).

4 Ordinal numerals

Ordinal numerals are created in a regular fashion by adding the suffix *-ste* or the suffix *-de*. The suffix *-ste* [stə] is added after the ordinal allomorph for *een* 'one', *eer-*, after *acht* 'eight', after the suffix *-tig* (*twintig-ste, dertig-ste*, etc.), and after the numerals *honderd, duizend, miljoen*, and *miljard*. In all other cases (after 2–7, 9–10 and numerals ending in these numerals), the suffix *-de* [də] is used.

(31) Dutch ordinals 1–10

1 eer-ste
2 twee-de
3 der-de
4 vier-de
5 vijf-de
6 zes-de
7 zeven-de
8 acht-ste
9 negen-de
10 tien-de

The final schwa of the ordinal suffixes is not the schwa of adjectival inflection; it is always present, also before singular neuter nouns in indefinite phrases where regular adjectives have no final schwa:

(32) een mooi huis 'a nice house'
een tweede huis / *een tweed huis 'a second house'

As pointed out by (Barbiers 2007), the word *eerste* is not a regular ordinal numeral, but a superlative form used as an ordinal. Dutch superlatives are formed by means of suffixation with *-ste*, and the word *eerste* can occur in the contexts in which superlatives occur. For instance, superlatives can be preceded by the intensifying prefix *aller-* 'most', as in *aller-mooi-st* 'most beautiful', and *aller-* can also precede *eerst*: *aller-eerst* 'very first'. On the other hand, the word *eer-ste* behaves as an ordinal in that the final schwa of the suffix is always present, even in indefinite neuter NPs, and thus differs from regular adjectives:

(33) een eerste / *eerst huis 'a first house'
een *mooie / mooi huis 'a nice house'

Ordinals for complex numerals are created by using the ordinal form of the last numeral only. Thus, ordinal formation may be qualified as a head operation (Hoeksema 1988). This applies to both Dutch and English.

(34) honderd en eerste / *een-de'101th'
honderd en derde / *drie-de '103rd'
honderd-twee-en-twintig-ste '122th'
drie-duizend-drie-honderd-ste '3300th'

Both in Dutch and English, irregular forms of digit ordinals such as *eer-ste* 'first' and *der-de* 'third' recur in the ordinals for complex numerals. Hence, we need the following two schemas for Dutch ordinals (that can be unified into one schema):

(35) a. $[Num^i + de/ste]^j_{Ord} \leftrightarrow [ORD\ [NUM^i]]^j$
b. $[Num^k + Ord^j]^l_{Ord} \leftrightarrow [ORD\ [NUM^k + NUM^j]]^l$
(where ORD is the semantic operator of Ordinality).

The Numeral base in (35a) is a word, simplex or complex (for instance, *tien, vjftien, twintig*,); in the default case it is words that form the bases of morphological operations. The interesting point is that the ordinal affix has formally scope over the last numeral only, as proven by the selection of irregular forms, whereas semantically it has scope over the whole complex numeral expression. The mismatch between form and meaning in the ordinal forms of complex numerals can be straighforwardly expressed by schema (35b), which refers to the NUM value (that is, a semantic property) of the ordinal. Such mismatches between form and meaning thus form an argument in favour of the use of constructional schemas.

5 Fraction numerals

In Dutch, as in many other European languages, cardinal and ordinal numerals are combined in the formation of names for fractions:

(36) drie-acht-ste '3/8'
twaalf-honderd-ste '12/100'

These fraction numerals can be seen as instantiations of the multiplication schema (20), since ordinals can function as measure nouns with fractional meaning.[2] Ordinals can be used to denote parts in combination with determiners as in:

(37) a. een acht-ste
a.SG eighth.SG.NEUT
'an eighth'

b. het acht-ste
the.SG eighth.SG.NEUT
'the eighth part'

c. vier vijfd-en van de toeristen
four fifth-PL of the tourists
'4/5 of the tourists'

These derived nouns have neuter gender, as shown by the choice of the determiner in (37b). As example (37c) shows, the fraction noun is not obligatorily a measure noun, as it can be pluralized. The following schema expresses the relevant generalization:

(38) $[\text{Ordinal}^{i}]^{j}{}_{\text{N ([+M]), [+ neut]}} \leftrightarrow [1/\text{NUM}^{i}]^{j}$
(where NUM^{i} is the arithmetical value of the corresponding numeral)

[2] Note that in French, fraction names are also substantivized ordinals, and have always plural forms, as in *trois huitièmes* '3/8'.

In this schema the feature [+M] is optional since such converted nouns can also be pluralized. This schema specifies the specific fractional meaning of converted ordinals.

Converted ordinals can also be preceded by definite determiners without receiving a fraction interpretation; in that case the noun is not necessarily a neuter noun:

(39) de achtste
the.SG.COMMON eighth
'the eighth (entity)'

Derde-n veroorzaakten de schade
Third-PL caused the damage
'A third party caused the damage'

When used as the basis of multiplication, the fraction interpretation of the converted ordinals is the only possible interpretation, and the noun is used as measure noun. The following subschema of (20) for fraction numbers expresses the fractional interpretation of multiplication structures with ordinals:

(40) Multiplication schema for fraction numerals
$[\text{Num}^i\ [\text{Ordinal}^j]_{\text{N, [+M]}}]^k_{\text{Num}} \leftrightarrow [\text{NUM}^i \text{ x } 1/\text{NUM}^j]^k$

Schema (40) is a unification of schemas (20) and (38), the schema for the conversion of ordinals into measure nouns with fraction meaning. Its properties are completely predictable, and hence it is not a construction of its own.[3]

Fraction numerals can be added to cardinal numerals by means of coordination, thus expressing addition. The coordination is usually asyndetic, as in:

(41) twee drie-vier-de '2¾'
zeventig drie-vier-de '70 ¾'

The numerals in (41) can thus be seen as instantiations of a schema that is a subpart of the general addition schema (24) for number names > 100, in which numerals with a higher value precede numerals with a lower value. In this subcase of cardinal numeral construction the conjunction *en* is, different from what is the case for numerals > 100, obligatorily absent. Once more, we observe that the construction of numerical expressions in Dutch reflects the syntax of Dutch, but is also subject to more specific restrictions as to the order of the numerals and the presence and form of linking elements for various subcategories of numerical expressions. This necessitates the assumption of specific schemas for numeral expressions.

The simplex word *half* '1/2' can also be used as part of a numerical expression. The word *half* is an adjective; the noun for 'half' in Dutch is *helft*. In complex numerals the word *half* can be preceded by the indefinite determiner *een*

[3] The fraction numeral for ½ is the non-derived lexical item *half* 'half'. It is only in mathematical contexts that one will use the regular expression *een-twee-de*.

[ən]. In such numerals, the fraction phrase *een half* 'a half' follows the numeral with the higher value. The conjunction *en* is optional, as illustrated here by the expression for 3½, for which two options are available:

(42) a. drie-[ən]-half (dri.jən)$_{\omega}$(hɑlf)$_{\omega}$
b. drie-[ɛn]-[ən]-half (dri)$_{\omega}$(ɛ.nən)$_{\omega}$(hɑlf)$_{\omega}$

The prosodic structure is predictable: each constituent forms a prosodic word of its own, except that the constituent -[ən]- has a schwa as its only vowel. Hence, it cannot form a prosodic word of its own, and will take the preceding word as its prosodic host. Therefore, in (42a) homorganic glide insertion takes place obligatorily before the schwa. In (42b), on the other hand, there is no vowel hiatus before the schwa.[4] In sum, the prosodic structure of such complex numeral expressions follows from general constraints on the construction of prosodic forms of Dutch words and phrases.

For complex fraction numerals of the type exemplified in (42) we need to assume a specific constructional idiom, that is, a pattern in which some positions are lexically filled whereas other are variable (Booij 2002). The constructional idiom, in which the non-initial constituents are specified lexically, will have the following form:[5]

(43) $[\text{Num}^{i} -([\text{ɛn}])_{\text{Conj}}\ [\text{ən}]\ [\text{half}]_{\text{A}}\ ]^{j}_{\text{Num}} \leftrightarrow [\text{NUM}^{i} + ½]^{j}$

This schema specifies how numerals with the fraction ½ can be formed productively, notwithstanding the idiosyncratic form of their fraction part.

6 Conclusions

In this article I have argued that a detailed analysis of the construction of Dutch numerical expressions throws light on the architecture of the grammar. Numerical expressions may be created by both syntactic and morphological means, and both play a role in the construction of Dutch numerals. The construction of Dutch numerals is partially a matter of conventionalized forms of syntactic coordination, subject to specific constraints. In the case of the numerals 21–99, the specific word order and the appearance of a linking element [ən] that derives historically from the conjunction *en* [ɛn], suggested the existence of a specific (morphological) schema for cardinal numeral compounds. These conventional restrictions on the use of the structural options provided by the syntactic constructions of Dutch reflect the Packing Strategy.

The analysis of ordinals showed how a mismatch between form and meaning in complex expressions can be accounted for by means of a schema. Moreover, the construction of numerical expressions with *half* illustrates the necessity of

[4] There is a vowel hiatus between the [i] and the [ɛ] which may be filled optionally by a homorganic glide in connected speech (Booij 1995).

[5] The numeral for 1½ is a lexicalized compound, *ander-half* 'lit. other half'.

more specific constructional idioms, schemas in which one or more of the constituents is lexically fixed.

Thus, the detailed analysis of the numeral system of Dutch provides clear evidence for an architecture of the grammar in which constructional schemas (both syntactic and morphological ones) of different degree of abstraction play a crucial role.

References

Barbiers, S. (2007): "Indefinite numerals ONE and MANY and the cause of ordinal suppletion". Lingua 117**,** 859–80.

Blevins, J. P. (2006): "Word-based morphology". Journal of Linguistics 42**,** 531–73.

Bochner, H. (1993): Simplicity in generative morphology. Berlin/New York: Mouton de Gruyter.

Booij, G. (1995): The phonology of Dutch. Oxford: Clarendon Press.

Booij, G. (2002): "Constructional idioms, morphology, and the Dutch lexicon". Journal of Germanic Linguistics 14, 301–27.

Booij, G. (2005): "Compounding and derivation: evidence for construction morphology". In: Dressler, W. U., D. Kastovsky, O. E. Pfeiffer & F. Rainer (eds.), Morphology and its demarcations. Amsterdam/Philadelphia: John Benjamins, 109–32.

Booij, G. (2007): "Construction morphology and the lexicon". In: Montermini, F., G. Boyé & N. Hathout (eds.) Selected proceedings of the 5th Décembrettes: Morphology in Toulouse. Somerville: Cascadilla Press, 34–44.

Booij, G. (2009): "Construction morphology and compounding". In: Lieber, R. & P. Stekauer (eds.) The handbook of compounding. Oxford: Oxford University Press, 201–16

Booij, G. & J. Rubach (1984): "Morphological and prosodic domains in Lexical Phonology". Phonology Yearbook 1**,** 1–27.

Briscoe, E., A. Copestake & V. de Paiva (eds.) (1993): "Inheritance, defaults and the lexicon". Cambridge: Cambridge University Press.

Bybee, J. L. (1995): "Regular morphology and the lexicon". Language and Cognitive Processes 10, 425–55.

Croft, W. (2001): Radical Construction Grammar. Syntactic theory in typological perspective. Oxford: Oxford University Press.

Culicover, P. W. & R. Jackendoff (2005): Simpler syntax. Oxford: Oxford University Press.

Culicover, P. W. & R. Jackendoff (2006): "The simpler syntax hypothesis". Trends in Cognitive Science 10, 413–8.

Evans, R. & G. Gazdar (1996): "DATR: a language for lexical knowledge representation". Computational Linguistics 22, 167–216.

Fried, M. & J.-O. Östman (2004): "Construction grammar: a thumbnail sketch". In: Fried, M. & J.-O. Östman (eds.) Construction Grammar in a cross-linguistic perspective. Amsterdam/Philadelphia: Benjamins.

Goldberg, A. (1995): Constructions. A Construction Grammar approach to argument structure. Chicago: Chicago University Press.

Goldberg, A. (2006): Constructions at work. The nature of generalization in language. Oxford: Oxford University Press.

Greenberg, J. H. (1978): "Generalizations about numeral systems". In: Greenberg, J. H. (ed.) Universals of human language. Vol. 3 Word structure. Stanford Calif.: Stanford University Press, 249–95.

Haeseryn, W., K. Romyn, G. Geerts, J. de Rooij & M. van den Toorn (1997): Algemene Neerlandse spraakkunst. Groningen/Deurne: Martinus Nijhoff/Wolters Plantyn.
Hoeksema, J. (1988): "Head-types in morpho-syntax". In: Booij, G. & J. v. Marle (eds.) Yearbook of Morphology 1988. Dordrecht: Foris, 123–38.
Hurford, J. R. (1975): The linguistic theory of numerals. Cambridge: Cambridge University Press.
Hurford, J. R. (1987): Language and number: the emergence of a cognitive system. Oxford: Basil Blackwell.
Hurford, J. R. (2003): "The interaction between numerals and nouns". In: Plank, F. (ed.) Noun phrase structure in the languages of Europe. Berlin/New York: Mouton de Gruyter, 561–620.
Hurford, J. R. (2007): "A performed practice explains a linguistic universal: Counting gives the Packing Strategy". Lingua 117, 773–83.
Jackendoff, R. (1975): "Semantic and morphological regularities in the lexicon". Language 51, 639–71.
Jackendoff, R. (2002): Foundations of language. Oxford: Oxford University Press.
Jackendoff, R. (2008): "Construction after construction and its theoretical challenge". Language 84, 8–28.
Kilbury, J., W. Petersen & C. Rumpf (2006): "Inheritance-based models of the lexicon". In: Wunderlich, D. (ed.) Advances in the theory of the lexicon. Berlin: Mouton de Gruyter, 429–80.
Langacker, R. (1987): Foundations of Cognitive Grammar: Theoretical prerequisites. Stanford, California: Stanford University Press.
Langacker, R. (1999): Grammar and conceptualization. Berlin: Mouton de Gruyter.
Lascarides, A. & A. Copestake (1999): "Default representation in constraint-based frameworks". Computational Linguistics 25**,** 55–106.
Östman, J.-O. (2005): "Construction discourse. A prolegomenon". In: Östman, J.-O. & M. Fried (eds.) Construction grammars. Cognitive grounding and theoretical extensions. Amsterdam/ Philadelphia: Benjamins, 121–44.
Paul, H. (1880 [3rd edition 1898]): Prinzipien der Sprachgeschichte. Halle: Max Niemeyer.
Riehemann, S. Z. (1998): "Type-based derivational morphology". Journal of Comparative Germanic Linguistics 2, 49–77.
Riehemann, S. Z. (2001): A constructional approach to idioms and word formation. Ph. D. dissertation Stanford University.
Sag, I. A. (2007): "Sign-based construction grammar. An informal synopsis". Manuscript. Stanford, Calif.: Stanford University.
Sag, I. A., T. Wasow & E. M. Bender (2003): Syntactic theory. A formal introduction. Stanford, Calif.: CSLI Publications.
Schultink, H. (1962): De morfologische valentie van het ongelede adjectief in modern Nederlands. Den Haag: Van Goor Zonen.
van Marle, J. (1985): On the paradigmatic dimension of morphological creativity. Dordrecht: Foris.
Wiese, H. (2007): "The co-evolution of number concepts and counting words". Lingua 117, 758–72.

Leiden, Netherlands Geert Booij

Leiden University Centre of Linguistics, Faculty of Humanities, P.O. Box 9515, NL-2300 RA Leiden, E-Mail: g.e.booij@hum.leidenuniv.nl