

# **Dutch morphosyntax and its interfaces**

**Habilitationsschrift**

**Dr. Marijke De Belder**

## Dutch morphosyntax and its interfaces

Dr. Marijke De Belder

Habilitationsschrift

Fachgebiet Niederländische Sprachwissenschaft

Eingereicht am 22. Dezember 2021

Institut für Niederlandistik, FK III

Carl von Ossietzky Universität Oldenburg

# Table of Contents

Declaration of sources	4
Acknowledgements	6
List of Abbreviations	7
<b>1. Introduction</b>	9
<b>Morphology-Syntax-Vocabulary</b>	
2. The root and nothing but the root: root primary compounding in Dutch	36
3. Linking elements: a case study on Dutch	65
4. Movement to D/d in compounds and phrases: referential compounding, -s possessives and title expressions in Dutch	100
5. One module, different levels of merge: AN(N) compounds in Dutch	120
6. Spectacular Splitters: Posture verbs, motion verbs and copulas in Dutch and other West-Germanic varieties	150
<b>Morphosyntax-Semantics-Vocabulary</b>	
7. The syntax-semantics interface in Distributed Morphology	193
<b>Morphosyntax-Phonology-Vocabulary</b>	
8. A split approach to the selection of allomorphs: vowel length alternating allomorphy in Dutch	211
9. Root allomorphy depends on head movement: support from Breton pluralization	249
10. The extravagant Dutch suffix <i>-ke</i> and its meandering through the interfaces	251
11. Category-specific syllable structures in Dutch: experimental evidence	273
<b>12. Conclusion</b>	310
Zusammenfassung	314
References	322

## Declaration of sources

Chapters of this dissertation have been accepted for publication and many of them have already appeared in press. The following bibliography makes this clear:

### Chapter 2:

De Belder, M. (2017). The root and nothing but the root: primary compounds in Dutch. *Syntax. A Journal of Theoretical, Experimental and Interdisciplinary Research* 20:2, pp.138-169.

### Chapter 3:

De Belder, M. (accepted) Linking elements: a case study on Dutch. Ackema P., Bendjaballah S., Bonet E. & Fábregas A. (eds.) *The Wiley Blackwell Companion to Morphology*. New York: John Wiley & Sons.

### Chapter 5:

De Belder, M., Van Koppen, M. (2016). One module, different levels of merge: AN(N) compounds in Dutch. *Studia Linguistica*, 70:1, pp. 1-33. DOI: 10.1111/stul.12040

### Chapter 7:

De Belder, M. (accepted) *The Syntax-Semantics Interface in Distributed Morphology*. To be published in Alexiadou, A., R. Kramer, A. Marantz & I. Oltra-Massuet (eds.) *The Cambridge Handbook of Distributed Morphology*. Cambridge: CUP.

### Chapter 8:

De Belder, M. (2020) A split approach to the selection of allomorphs: vowel length alternating allomorphy in Dutch. *Glossa: a journal of general linguistics* 5(1), 42. DOI: doi.org/10.5334/gjgl.899

### Chapter 9:

De Belder, M. (2018) Root allomorphy depends on head movement: support from Breton pluralisation. *Snippets* 34. DOI: 10.7358/snip-2018-034-beld

### Chapter 10:

De Belder, M. (In press) The extravagant Dutch affix -ke and its meandering through the interfaces. In: Matthias Eitelmann & Dagmar Haumann (eds.) *Extravagant Morphology: Studies in rule-bending, pattern-extending and theory-challenging morphology*. Studies in Language Companion Series. John Benjamins Publishing Company.

An adapted version of chapter 4 will appear in a Festschrift:

De Belder, Marijke (to appear) In: Adger, D., L. Martí, I. Roy and L. Stockall (eds.) *Festschrift for Hagit Borer*.

As can be concluded from the bibliography, one section has been co-authored with Prof. dr. Marjo Van Koppen, Professor of Dutch at the Meertens Instituut in Amsterdam and Utrecht University.

Furthermore, section 3 in the introduction (i.e. chapter 1) is partially taken from the following co-authored contribution with Dr. Jan Don from the University of Amsterdam (UvA):

De Belder, M. & J. Don (to appear) Distributed Morphology: An oratio pro domo. *Nederlandse Taalkunde*.

Apart from this section, De Belder & Don (to appear) is not included in the dissertation.

Finally, chapter 11 has been co-authored with Prof. Dr. Esther Ruigendijk, professor of Dutch at the Carl von Ossietzky Universität Oldenburg.

## Acknowledgements

I would like to express my deepest, heartfelt appreciation and gratitude to Prof Dr. Esther Ruigendijk for her support and encouragement. I would also like to thank the department of Dutch at the University of Oldenburg for the support.

For support in one way or another in recent years I would like to thank my sister Sarah De Belder and Dr. Ankelien Schippers, Dr. Bénédicte Grandon, Kerstin Bohne, Dr. Margreet Vogelzang and Andreas Hiemstra, Anja Glaser, Anja Hiller, Anna Schill, Prof. Dr. Anne Breitbarth, Prof. Dr. Artemis Alexiadou, Drs. Carla Broeder, Dr. Cora Cavigani-Pots, Prof. Dr. Dagmar Haumann, Prof. Dr. Dalina Kallulli, Prof. Dr. Dany Jaspers, Demetris Karayiannis, Dr. Dorothee Poggel, Prof. Dr. Eulàlia Bonet, Dr. Eva Zimmermann, Drs. Foekje Reitsma, Prof. Dr. Em. Geert Booij, Prof. Dr. Em. Gerd Hentschel, Dr. Gertjan Postma, Prof. Dr. Guido Vanden Wyngaerd, Prof. Dr. Hagit Borer, Dr. Haimo Stiemer, Drs. Hans Beelen, Prof. Dr. Harry van der Hulst, Prof. Dr. Heidi Harley, Dr. Heidi Klockmann, Ioanna Papadopoulou, Prof. Dr. Johan Rooryck, Prof. Dr. Jörg Peters, Juan Navarro, Dr. Karen De Clercq, Dr. Karolin Bubke, Katlijn Van Audenaerde, Katrin Hoppe, Katrin Naert, Prof. Dr. Kerstin Schill, Prof. Dr. Kleanthes K. Grohmann, Prof. Dr. Em. Liliane Haegeman, Dr. Maike Engelhardt, Maike Rudat, Prof. Dr. Marc van Oostendorp, Mr. Marcello Jansen, Prof. Dr. Marjo van Koppen, Dr. Matthias Eitelmann, Prof. Dr. Matthias Hüning, Prof. Dr. Nanna Fuhrhop, Dr. Noam Faust, Prof. Dr. Norbert Corver, Dr. Pavel Caha, Prof. Dr. Petra Hendriks, Petra Kunda, Prof. Dr. Phoevos Panagiotidis, Prof. Dr. Ralf Grüttemeier, Prof. Dr. Roberta d'Alessandro, Sabine Lohwasser, Seith Onyango, Snoes, Stephan Matthias, Dr. Victoria Marin.

## List of abbreviations

1, 2, 3	first, second and third person
$\pi$	phonological root index
$\sigma$	syllable
$\checkmark$	root, lexeme
$\emptyset$	empty set / zero exponent
A	adjective
ADV	adverb
ALL	allomorph
C	consonant
CARD	cardinal number
D	determiner
DIM	diminutive
DM	Distributed Morphology
EC	epenthetic consonant
F	feminine
FEM	feminine
HON	honorific
IMP	imperative
INDEF	indefinite
INF	infinitive
INFL	inflection
L	allomorph with a long vowel
LE	linking element
LF	logical form
LP	linking phoneme
m-	modificational
MASC	masculine
MONOPH.	monophthong
n	absolute number
(little) n	nominalizing head
N	noun
NCM	nominal compound marker / nominal class marker
NUM	number marking
OBSTR.	obstruent
OT	Optimality Theory
PF	phonological form
PL	plural
r-	referential
REFL	reflexive pronoun
RPC	root primary compound
S	allomorph with a short vowel
SG	singular
SON.	sonorant

Spec	specifier
(little) v	head that introduces an external argument into the verbal structure
V	verb / vowel
W	phonological word
XP	phrase of the type X

# 1 Introduction

## 1.1 How to organise which interfaces

The present work discusses Dutch morphology and its interfaces. Morphology has various interface relations. Firstly, morphology of course has an interface relation with the language's vocabulary<sup>1</sup>. Morphology builds words and the word's morphemes are realised by vocabulary items (i.e. roots and affixes). Secondly, given the obvious fact that sentences consist of words, there must be a close interface relation between whichever module that builds words and the one that builds sentences. Thirdly, the abstract output of morphology is the input of word phonology. Fourthly, words mean something. They may contribute lexical semantics and functional semantics to the sentence. In fact, they may even interface with pragmatics if they contain, for example, honorific affixes.

The linguist studying morphology and its interfaces faces two important challenges. Firstly, data do not come labelled, thus the first problem is often to determine for which module or interface between modules the empirical problem may be relevant. For example, is allomorphy selection a strict problem of vocabulary (i.e. vocabulary has several variations of an exponent and needs to organise them), is it an effect of the way morphosyntactic structure *interfaces* with vocabulary (i.e. one of the allomorphs may be marked to realise a specific morphosyntactic structure) or is it a phonological problem (e.g. one allomorph may simply be phonologically more optimal in the context). The first research question of the present bundle is therefore the question which module or which interface is the adequate locus for certain seemingly hybrid data. The analysis of such data often have been under debate for decennia. For example, the present dissertation presents compounds that clearly contain a syntactic phrase or a fragment of syntactic functional structure. One could of course wonder what such compounds teach us about compound formation. More generally, however, one could ask what we learn about the relation between morphology and syntax from such data. Another example is syntactic polysemy. If a verb can be used both as a lexical verb and as a copula, is this then a complication at the syntax, at vocabulary or perhaps both? What do such data tell us about these modules and how they interface?

The second challenge is that the data typically do not show a neat ordered interaction between two modules. Instead, all modules seem to interact with one another simultaneously. Consider the following example. It will become clear that the insertion of a specific epenthetic consonant in Dutch is determined by a simultaneous occurrence of a specific morphosyntactic, lexical and phonological context and it is blocked by an honorific (i.e. a specific pragmatic) context. Theoretically, such a state of affairs is not expected. One expects that during the derivation of a structure the output of one module (e.g. syntax) functions as the input of another module (e.g. phonology or semantics) and one expects these modules to be organised in a specific order such that output can become input. As such, the default hypothesis would be that one finds neat output-input interface interactions. I have found, however, that this is rarely the case empirically. There is thus a tension between the complexity suggested by the data and the fact that we can reasonably assume that the model must be a strictly organised process. The second theme of the dissertation is

---

<sup>1</sup> The term 'Vocabulary' refers to what is traditionally known as the mental lexicon. Given that the present work adopt the framework of Distributed Morphology, I use its terms to avoid any confusion.

thus the question how the Y-model (see section 2 in this chapter) and, more specifically, the interfaces in the Y-model should be organised to capture the data. I thereby take into account theoretical considerations that address the question what we find a desirable organisation on principled grounds and technical considerations, guaranteeing that at the end of the day we have the tools to derive the surface representation.

I thus address both challenges. I present empirical domains that are relevant for our understanding of which module and which interfaces are relevant to which phenomenon and I discuss theoretical considerations and technical solutions to organise an adequate interface model. Nearly all contributions discuss Dutch morphosyntax. Two exceptions are the brief contribution in chapter 9 on Breton and the more general theoretical contribution in chapter 7 that discusses relevant examples from various languages.

The remainder of this introduction is structured as follows. In the next section I present the framework in which the work was written, i.e. Distributed Morphology and the consequences this theoretical choice has for the approach to interfaces. In section 3 I argue that morphology does not interface with syntax, as morphology is syntax. In section 4 I discuss the morphosyntax-lexicon interface, section 5 addresses the morphosyntax-semantics interface and section 6 contains information on the morphosyntax-phonology interface. Section 7 contains an overview of the chapters in this dissertation and comments on how they contribute to its overarching theme.

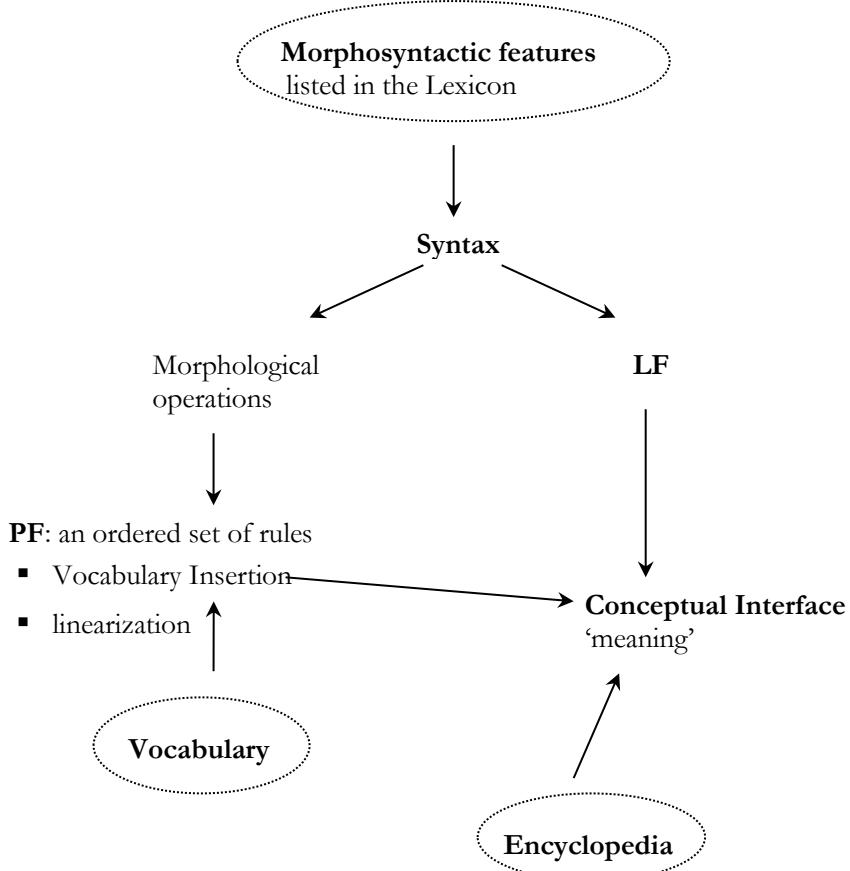
## 1.2 Distributed Morphology

The present work is embedded in the framework of Distributed Morphology and I thus adopt the Y-model as proposed by Halle and Marantz (1993), Harley and Noyer (1999) and references therein. It is illustrated in (1) (which draws heavily upon the picture in Harley and Noyer 1999:4 and see De Belder 2011a:20).<sup>2</sup>

---

<sup>2</sup> The dotted ovals indicate lists (as opposed to operations).

(1)



In this model, syntax operates on morphosyntactic features and roots. The morphosyntactic features are functional ones. They are provided by UG and stored in an inventory called the Lexicon.<sup>3</sup> Examples of features are [sg], [1<sup>st</sup>], [past], [focus], [distal], [Voice], etc. All inflectional features and all features relevant to derivational word formation also belong to this set of morphosyntactic features. As such, inflection and derivation are part of syntax. Indeed, the idea that morphology is syntax or ‘syntax all the way down’ is a defining characteristic of Distributed Morphology (see section 3 in this chapter for discussion). The complete feature set on one terminal node is called a morpheme. The morpheme is thus an abstract notion and should not be confused in this model with its phonological realisation, called an exponent.

Open-class lexical items (i.e. roots) can be defined in various ways (see section 5 in this chapter for details), but in contemporary contribution it is usually defined by an index. The precise nature of the root index is under debate: it may be a fully abstract index or it may be a partially abstract phonological index (see Borer 2013a, Harley 2014). To express that the index is of an arbitrary, abstract nature, the convention is to refer to it with an arbitrarily chosen number. For example, the root index 317 may refer to the root vocabulary item *cat*. Given that the root index is part of the Numeration from which syntax selects elements to build its structure, the root indices must be listed in the Lexicon as well. As a result, the Lexicon then contains both innate features and learned root

---

<sup>3</sup> The term Lexicon should thus not be confused with what is known as the mental lexicon. The inventory of the language’s lemmas is called the Vocabulary.

indices. I discuss alternative views below, according to which the Lexicon would only contain innate syntactic features.

Note that the root node contains at most an index. It does not contain any morphosyntactic features, not even categorial ones. This has far-reaching consequences for the way syntactic structure is built and interpreted (see sections 4 and 5 in this chapter for details). The lexical core of the structure does not project any syntactic properties into the structure. All properties of the derivation are derived through the merger of functional heads. Functional heads are nodes that contain morphosyntactic features which are given in the Lexicon. Examples of such heads are Tense, D, n, V, Number,... The number and nature of functional heads merging on top of a root will determine whether that root is interpreted as a noun, a verb or an adjective. If, for example, it is to be interpreted as a verb, the functional heads also determine the aspectual properties, the nature and number of the arguments and their theta-roles, etc. The root *work*, for example, will become a verb under such verbal functional heads. It will, in contrast, become a noun under nominal structure. This could be, for example, number and a determiner, but also a nominal derivational head such as the nominal head realised by the affix *-er* as in *worker*.

Such derivational heads, which will be realised by derivational affixes, are called ‘little heads’. A nominal derivational head is little n, a verbal one is little v, etc. The names hang together with the convention not to capitalise the letters in the syntactic tree (nP, vP, ...). There is a debate whether one should postulate productive zero affixes to assign a category to a root (see De Belder 2014 for discussion). Distributed Morphology assumes such heads across the board, even when there is no direct empirical evidence to do so. For example, the root *cat* will merge with little n prior to merging with the rest of the nominal functional sequence. Structurally, the noun *cat* therefore does not differ from the noun *worker*. This approach has often been criticised and Borer (2013a) argues, in my opinion convincingly, that conceptually and technically, there is no need to assume such productive zero little heads. I have always followed Borer’s Exo-Skeletal approach in this respect and I do not assume productive zero little heads. In my work, the noun *cat* and the noun *worker* thus do differ structurally: the noun *cat* is nominalised solely by its nominal functional heads (gender, number, D, ...), the noun *worker* is nominalised by little n (realised by *-er*) and its nominal functional heads.

Syntactic structures are sent to PF and LF. As such, the Distributed Morphology model is a Y-model as proposed in the Minimalist Program (Chomsky 1995). As the structure is sent to LF, the morphosyntactic features present at Spell-Out are interpreted compositionally, regardless of further operations on the PF side, unless one assumes an additional interface between PF and LF, which may or may not be done in Distributed Morphology (see section 5 in this chapter for detailed discussion, but note for now that the model depicted in (1) does assume such an interface: there is a line that directly connects PF to LF).

Distributed Morphology adopts the familiar idea of generative syntax that Syntax spells out phases or cyclic domains (Chomsky 2001, Marantz 2007, Embick 2010). In other words, Syntax does not wait until a complete clause is finished to ship off its structure. It is rather assumed that smaller units are spelled out. The exact size of such units is under debate. Relevant for this dissertation is that authors in Distributed Morphology often assume that such units can be rather small, i.e. smaller than the morphological word. Little heads are often seen as phase heads, i.e. they trigger spell-out, in Distributed Morphology. As a consequence, a word that contains several derivational affixes may technically consist of several phases. This theoretical proposal makes important empirical predictions: given that phases are the units that are spelled out to the phonological and semantic interfaces, they are seen as units at these modules. Units of phonological and semantic irregularities

(i.e. idiomatic meaning) should thus instantiate a single phase to guarantee the required locality to derive the irregularity (Marantz 2007, Embick 2010 and see e.g. De Belder, Faust and Lampitelli 2014). For example, an allomorph (i.e. an instance of phonological irregularity) should be spelled out together with its determining context. Even though there is debate on which heads exactly are phase heads, there is a consensus in Distributed Morphology on the fact that irregular patterns and the contexts determining these patterns should be visible together at the interfaces.

Before PF proper, so-called morphological operations on the syntactic structure may take place, such as fission and impoverishment (see Bonet 1991, Halle & Marantz 1993 and Noyer 1997 for a detailed discussion). These operations may alter the feature set on a terminal node. It is perhaps confusing that these operations are called morphological operations as they are not directly to be associated with what is understood to be morphology in traditional terms. Indeed, it is not the role of these morphological operations to build words. They do alter the featural content of the morpheme (i.e. the tree's terminal node), though. Note that these operations do not affect the semantic interpretation of the terminal node, as the syntactic structure had been shipped already to the semantic component in the Y-model, unless, of course, one assumes a direct interface between PF and LF (see below and see section 5 in this chapter).

PF is conceived of as an ordered set of rules (Embick 2007). Amongst other things, it is the domain of Vocabulary Insertion and linearization. The phonological exponent of a vocabulary item is thus inserted post-syntactically, a process known as late insertion. Vocabulary items are stored in a learned list which is called Vocabulary. Vocabulary insertion is an operation which searches Vocabulary for an appropriate vocabulary item and inserts it into the structure.

A vocabulary item consists of a phonological exponent and information on the feature set it can realize. Examples are given in (2)-(3).

- (2) /ðæt/ ↔ [definite, distal, singular]
- (3) /kæt/ ↔ √317

Example (2) shows a demonstrative pronoun, example (3) shows a root. The left-hand part of the vocabulary item specifies its phonological form. This is called the phonological exponent. The righthand part specifies the context in which it can be inserted. As can be seen, functional contexts are defined by syntactic feature sets, roots, as mentioned above, are defined by an index. The index expresses that at syntax an abstract index has merged that refers to the root *cat*, at vocabulary a vocabulary item will be stored that matches the correct exponent (e.g. /kæt/) to this index. Roots are represented by means of the symbol √. Note that vocabulary insertion does not match meaning to the index. After all, the semantic side of the structure is not dealt with at the set of operations at the PF-side of the model. The context of insertion may be enriched with extra specifications regarding the appropriate context for insertion (see sections 5 and 6 in this chapter).

The features for which the *vocabulary item* is specified only come into play at PF as vocabulary items are inserted post-syntactically. The exact set of features for which the vocabulary item is marked is thus irrelevant for syntax and LF. For this reason no one-to-one correspondence is required between the feature set of a syntactic node and that of the vocabulary item which will realise this node. All that is needed is that the vocabulary item is the best possible candidate to realise this node. Insertion is regulated by the Subset Principle, which states that all vocabulary items of which the features are a subset of the terminal node's feature set could be inserted and the closest match wins. In other words, late insertion allows in principle for certain mismatches

between the feature set of the terminal node and that of the vocabulary item. This is an important advantage over early insertion models for which it is difficult to implement such a mismatch. As such, it motivates the postulation of late insertion. An example of an analysis that illustrates the descriptive advantage of such a model is given in chapter 10.

The conceptual interface receives input from three different modules in order to assign a meaning to the derivation. Firstly, it interprets the output of LF. LF is the interface which takes the syntactic structure as its input. It interprets the morphosyntactic features, such as ‘plural’ or ‘definite’, compositionally. This ensures that all interpretable features which were present in the syntax will be interpreted. Secondly, it can access information from Encyclopedia. Encyclopedia is a learned list which matches constituents and the vocabulary items that may realise them to an interpretation. It contains the information that a VP realised as *kick the bucket* should be understood as ‘die’, but it has also stored the meaning of the word *cat*. In Encyclopedia any meaning can in principle be attached to any structure. However, there is one important restriction. The compositional meaning which is computed at LF on the basis of morphosyntactic features can never be overridden (McGinnis 2002). Idiomatic VPs get the same aspectual interpretation as non-idiomatic VPs with the same structure. For example, a VP which contains the root *hang* and a DP-object marked for singular number is telic in English, as opposed to one which has a mass DP-object, which is atelic, as can be made explicit by the temporal adjunct indicating a time limit. The observation holds true both for non-idiomatic examples (4)-(5) and idiomatic ones (6)-(7) (examples are taken from McGinnis 2002:668):

- |  |                         |
|--|-------------------------|
| (4) Hermione hung a picture in five minutes. | [telic, non-idiomatic]  |
| (5) Hermione hung laundry <*in an hour>.     | [atelic, non-idiomatic] |
|  |                         |
| (6) Hermione hung a left in five minutes.    | [telic, non-idiomatic]  |
| (7) Hermione hung fire <*in a week>.         | [atelic, non-idiomatic] |

What these examples show is that the semantics associated with syntactic structure is retained even in idiomatic interpretations. Stored lexical information cannot override the compositional meaning of syntactic features. The compositional semantics of the morphosyntactic features which were merged in syntax is thus preserved throughout the derivation. It is a reasonable assumption that what LF computes therefore serves as input to the interpretational interface and this input is not subject to lexical knowledge. Note that in the literature the term LF is used ambiguously; it sometimes refers to the smaller module that interprets the features compositionally and it is sometimes used as a synonym for the conceptual interface.

In the model in (1) one can see an arrow from PF to the conceptual interface. When this model was originally drawn and published in 1999, the necessity of this arrow was clear. At syntax, the feature [Root] would merge at the root node and only at PF a specific root vocabulary item would be inserted. For example, only at PF would we learn that the root in the DP refers to a *cat* and not a *dog*. This output was then needed at the conceptual interface, otherwise the meaning part of the language would not be able to determine whether the structure semantically refers to *cats* or *dogs*. If one assumes, however, the presence of a root index at syntax, which has become the consensus since Harley (2014), this interface relation is no longer necessary. For example, Syntax would contain the index  $\sqrt{317}$  and subsequently, PF is able to determine the appropriate exponent to realise the node and the conceptual interface knows that the node refers to man’s second-best furry

friend. The root's index guarantees the constancy of the root node at both LF and PF, no direct relation between PF and LF is needed anymore if one assumes indices. I come back to this issue in section 5 in this chapter.

In sum, Distributed Morphology assumes the Y-model of the Minimalist Program: syntax spells out cyclic domains which are shipped off to PF and LF. Nevertheless, DM has some unique characteristics as well. Most notably, it postulates a syntax which exclusively operates with abstract features. Consequently, the PF side of the model is richer. It is conceived of as an ordered set of operations, one of which is Vocabulary Insertion. Vocabulary items are thus inserted post-syntactically. An important consequence of late insertion is that there does not need to be a one-to-one correspondence between the features in syntax and the features of the vocabulary items realising them. It further does not assume a specific module responsible for word formation. Morphology is distributed over syntax and a set of post-syntactic operations.

The structure of this Y-model has immediate consequences for several interfaces. Firstly, as I have noted, the framework only assumes a single structure-generating module called Syntax. Indeed, the generation of words and sentences takes places at the same module. In other words, morphology *is* syntax and I will therefore call all structure, words or sentences, morphosyntactic structure in this work. It is thus technically inaccurate to speak about the morphology-syntax interface in this framework: a module does not interface with itself. This, however, does not make the relation between syntactic structure and morphological structure less interesting, as will become clear in later chapters. Any morphologist has to grasp the relation and the divide between sentential structure and structure that defines wordhood, regardless their stance on how many modules are involved in the account.

Secondly, the relation between syntax and the Vocabulary in Distributed Morphology is strongly characterised by the root-based nature of the framework. As we have seen, the framework does not assume categorial features for open-class lexical items in the lexicon. This hypothesis has far-reaching consequences for the vocabulary-morphosyntax interface (and the semantics-morphosyntax interface). The present work contributes to our understanding of these consequences.

Thirdly, there is no consensus in Distributed Morphology on how morphology and phonology interface in the module. Given that the module adopts the general Y-structure according to which syntax spells out to PF and LF, the basic order of the interfaces is agreed upon. However, the specific technicalities of the interaction between morphology and phonology are still under debate. The present work adds to this debate by discussing case studies that contribute to our understanding of the complexity of the data and it offers theoretical proposals and technical solutions.

Fourthly, the interface between morphology and semantics is rarely the main topic of work in Distributed Morphology, but it is interesting given the field's specific view on vocabulary. If root vocabulary items are not specified for categories or any other features, where does categorial meaning come from? Different proposals have resulted in very different assumptions about morphosyntactic structure. The present work contains a critical overview of the state of the art of the framework in this respect.

In short, the present dissertation is embedded in Distributed Morphology. The way the model is conceived in this framework immediately defines how certain interfaces are to be understood, but it also determines the open questions. The present proposals contribute to this framework.

### 1.3 Morphology is syntax

It is an important and legitimate question whether words and sentences are built in the same structure-generating module in the human cognition or not and it is a question that has been central in the work of many morphologists of the past century. In this dissertation it deserves a discussion: if one wants to study morphology and its interface, the very first question whether morphology is syntax or whether morphology rather interfaces with syntax is not a detail. As I have pointed out above, I do adopt the idea that morphology is syntax. This is partially justified by the fact that I adopt the framework of Distributed Morphology, but I will further justify this choice by means of the discussion below.

As pointed out by Borer (2001:151) the question whether morphology and syntax are one and the same module is an empirical one: can we find morphological phenomena that cannot be captured by independently motivated syntactic principles? After all, in principle, assuming a single module is the simpler assumption and more than one module should not be assumed without reason to do so by Ockham's razor (Marantz 1997). What do we gain in terms of adequacy by assuming an independent word-building module? Observationally, it is not hard to come up with possible empirical candidates that would justify some caution as to assume a single module. Regardless the position one takes on the number of structure-generating modules, they need to be addressed. Proponents of the Lexicalist Hypothesis (i.e. two separate modules) have to argue the issues are pressing enough to justify the assumption of the extra module, proponents of the "Syntax all the way down" hypothesis (i.e. one module) have to account for the phenomena referring to syntactic principles. In what follows I describe relevant observations known to the field and I briefly comment on them (see also Bruening 2018 for more arguments and references).

A first serious concern is the issue of wordhood. Words are psychologically real. Speakers feel that sentences consist of words and speakers may have judgments on whether a word 'exists' or not. The term 'neologism' has no syntactic counterpart: we have no term for sentences one never has heard before. There is no syntactic counterpart of the dictionary either. The phonological word is also a relevant domain within the prosody hierarchy: specific phonological processes may refer to the phonological word or word boundaries. The word is also a morphological domain. Allomorphy, for example, is triggered by elements within the boundaries of the word (see chapter 9). In sum, a framework that only assumes a single structure-generating module still has to recognise the observational distinction between words and sentences. In Distributed Morphology, wordhood is derived by incorporation (see Mithun 1984 and Harley 2009), which is basically head movement, an independently accepted syntactic principle. The principled distinction between words and sentences is thus recognised and captured in the framework by the fact that a specific syntactic operation defines wordhood.<sup>4</sup> At the conceptual interface, incorporated structure will then be matched to stored structures recognised as words.

Two older arguments that were traditionally seen as arguments in favour of the Lexicalist Hypothesis are the observation that derivation may feed inflection but not the other way around (Chomsky 1970, see a.o. Beard for discussion 2001) and the Righthand Head Rule. I discuss them in turn below.

Derivational affixes are found closer to the stem than inflectional affixes. Derived words get inflected, inflected words are not subject to derivation. *Neighborhoods* is fine, *\*neighborhood* is not.

---

<sup>4</sup> In Distributed Morphology, compounds are usually drawn as incorporating structures, but morphological derivations are not necessarily drawn as such. This is an issue that perhaps should get more attention in the field.

A weak Lexicalist model entails that derivation takes place in a designated morphological module, whereas inflection would take place in syntax. It is further entailed by the Lexicalist Hypothesis that morphology feeds syntax: first words are created and these words enter the enumeration for the syntactic derivation. The fact that morphological derivation then takes place prior to inflection follows immediately from such a model. However, we know that strict ordering may take place in syntax as well. If one assumes a functional sequence with a strict order of functional heads (Cinque 1999), the inherent order of a syntactic derivation follows. As such, independently syntactic principles suffice to account for the observed order of word derivation and inflection. One simply has to assume that derivational heads precede inflectional heads in the functional spine. The inherent order of word derivation and inflection therefore does not necessitate the assumption of two separate modules.

The righthand head rule was argued to be a morphology-specific principle, supporting the Lexicalist Hypothesis. The idea was that a head was defined in morphology by virtue of being ‘at the righthand side’, i.e. by a linear principle, a syntactic head was defined by Williams as ‘the subphrase having all the same category features, but one bar fewer’ (Williams 1981:248). Looking back, I believe that in this year of writing we can confidently say that, despite its fame, the argument was a rather weak one. Firstly, the rule is observationally inadequate, as was already noticed by Williams in the original article. Prefixes may be the head of a word, the lefthand part of a compound may be the head, as in the English word *push up* (as noted by Williams himself) or as in the Dutch compound *de regering-Biden* ‘the Biden administration’, and the role of exocentric compounds in the argument is unclear, as was admitted in the original work.

Secondly, we should also recognise that the righthand head rule is anglocentric. It works fine as a rule-of-thumb for students studying Germanic languages, but in Romance languages for most derivational processes the head is situated at the righthand side (i.e. it is the suffix, as in *classement* ‘classification’), whereas for compounds it is typically to be found at the lefthand side (as in *rouge-à-lèvres* ‘red-for-lips, i.e. lipstick’). In French morphology, headedness thus does not depend on a simple linear principle. An argument that would substantiate something as fundamental as the number of modules in human cognition is not supposed to fall apart by something as simple as looking at the next well-described Indo-European language.

Thirdly, theoretically, it is not clear why Williams insisted on the distinction between morphological and syntactic heads in the first place (other than it serving as support for the Lexicalist Hypothesis). He actually gives exactly the same definition for syntactic and morphological heads. This single definition is cited here (the (8) is part of the citation):

The notion “head of a unit” is familiar from both traditional and recent syntax and morphology. In general, the head of X has the same properties (distribution, etc.) as X. We may then define the general notion “head” as shown in (8):

(8) If both X and the head of X are eligible members of category C, then  $X \in C \equiv^5 \text{head of } X \in C$  (Williams 1981:247)

---

<sup>5</sup> From the context I take it that  $\equiv$  denotes ‘identical’ in Williams’ text, which is one of the possible meanings of the symbol.

The fact that the head is actually the same concept in morphology and syntax could actually be taken as a strong argument in favour of a single module. Consider that Williams' definition of the head in syntax, i.e. 'the subphrase having all the same category features, but one bar fewer' works fine for morphological structures as well (consider: 'the head of a morphological structure is the node having all the same category features, but one bar fewer'). In more contemporary terms, I would say that the head of a structure (i.e. any structure) is the node that projects.

In Williams (2007) the distinction between syntactic and morphological headedness is reduced to the claim the morphology is head-final and syntax is head-initial in English. As pointed out by Bruening (2018) this distinction is rather superficial and counterexamples in both directions can be found easily. This language-specific argument is simply not strong enough to necessitate the assumption of a separate module.

An observation that goes back to Chomsky (1970) is that morphological processes may differ qua semantic transparency, regularity and productivity from syntactic ones. For example, English gerund formation (a syntactic transformation) is regular, transparent and productive, whereas English nominalisation (a morphological derivation) is not necessarily so. It is commonly assumed that such processes differ from one another and one way to define the distinction is by assigning the more regular process to syntax (or 'transformations' in older terms) and the irregular processes to the lexicon or a designated module for word-building.<sup>6</sup>

The argument still resonates in more contemporary approaches (see e.g. Newmeyer 2009, Ackema and Neeleman 2004). However, the idea faces considerable empirical problems, the most obvious one being that idiomticity is not restricted to words, but well attested for idiomatic constituents and sentences as well. More generally, this shows that idiomticity should be derivable in syntax as well. Idiomticity for words can then be derived by more general independently motivated principles known from syntax. More fine-grained empirical problems in this respect are presented in chapter 5 (and see also Bruening 2018). Section 3 in chapter 2 adds a note to the discussion by making the more general point that not every word that looks morphologically complex is also morphologically complex. After all, when discussing words that are supposed to illustrate highly irregular morphological processes, one should make sure to guarantee the first step of the argumentation, i.e. to show that the data under discussion are indeed instances of morphologically complex structures.

Having addressed classic arguments in favour of two modules, I will now turn to more recent arguments on the difference between words and phrases. Ackema & Neeleman (2004) observe that stranding is not possible for words in languages such as English. They use this as an argument against the idea that words are built in the syntax, the general reasoning being that 'if x is possible in phrases, then why is it not possible in words if one assumes a single module'? Here is an example of their central observation (Ackema & Neeleman 2004: 332):

- (8) a. the centre of a prosperous medieval city in Northern Italy  
b.\*the city<sub>i</sub> centre of a prosperous medieval t<sub>i</sub> in Northern Italy

The argument is straightforward: if word-formation involves head movement, why is head-movement excluded in the above example? Does Distributed Morphology not predict (8)b to be a

---

<sup>6</sup> There is, intriguingly enough, no consensus in the field on whether Chomsky (1970) should be read as the starting point for the Lexicalist Hypothesis or for roots (see e.g. Embick 2021). In any case, historically, it did result in the exploration of both ideas.

grammatical structure of English? It is clear that nothing is wrong with the compound *city centre*. The culprits are the stranded adjectival and prepositional modifiers that now suddenly lack a head. Clearly, English does not allow a head to be moved out of a full DP, ‘stranding’ any modifying material.

Before answering this empirical challenge, let us first make matters worse and look at a second empirical problem that Ackema & Neeleman 2004 raise for the word-formation-as-syntax-theory.

- (9) a. \* driver truck
- b. he drives a truck
- c. truckdriver

What this and similar examples show, according to Ackema & Neeleman (2004), is that complex nouns built from a verb cannot license the internal argument in the same way as the verb can. This is, however, according to Ackema & Neeleman (2004), what would be expected on the assumption that words are built in the syntax. Their reasoning is based on the structure in (10) that should underlie (9)c in a DM-type theory:

- (10) \*[NP [N [V drive] er] [VP tV [NP a truck]]]

Crucial is the idea that the structure would start out with a VP. However, we may assume alternatively (following Harley 2009) that (9) has a different underlying structure:

- (11) [n -er [vP √DRIVE [nP truck] ] ]

In (11) the complement *truck* is a complement of the category-neutral root DRIVE, rather than a complement of the noun. That is, before the root DRIVE is nominalized, it is merged with its complement. Next, the root + complement are merged with the nominal head *-er*. This way of analyzing (9)c does not immediately imply that (9)a would be grammatical. What prevents (9)a under this analysis is the claim that internal arguments are internal arguments of the root, rather than internal arguments of the verb (or the noun) (and see Borer 2011 who even argues against an analysis according to which *truck* would technically be an internal argument within the compound).

Let us now go back to the example in (8). Again, we would argue that a compound arises from a structure in which the root of the head of the nominal compound (here *centre*) merges with the root of the left-hand member (here *city*). Now, does DM predict that the structure in (8)b is derivable? Well, certainly not. The hypothetical derivation presents several hurdles that cannot be taken. The obvious first problem is the head movement constraint (Travis 1984:131) which states that no governing head position can be skipped. Long head movement is thus successive-cyclic head movement. The noun *city* would thus pass the DP-layers, the preposition and perhaps the adjectives. Within DM, the head incorporation would create word-hood: the fragment *city centre of a medieval* would roll up into one large head incorporation resulting in one big head. It would no longer be a sentence.

On its way, the head would encounter a D-head, adding referential semantics to ‘city’. This would result in a semantic contradiction: the non-head of the compound would be both referential (due to the D-layer) and non-referential. It is commonly assumed in Distributed Morphology that D-layers are illicit for non-heads in modificational compounds (see chapter 3 for discussion), for

semantic reasons. Deleting the semantics of the D-layer is not an option. Meaning is assumed to be built in a compositional and monotone fashion. Even if one would not adopt the successive-cyclic head movement, the problem would remain that ‘city’ would be both referential and non-referential simultaneously.

Furthermore, there is the more general hypothesis that a single head cannot host two words (Roberts 1997). ‘City’ would be a word in its base position: the root has its nominal functional structure and thus acts as a noun in the DP. In its landing position it would function as a part of a word. This is illicit due to Roberts’ more general constraint that words cannot incorporate into words to become a part of a word.

More generally, the idea that morphology is syntax does not do away with the idea that a word is a reality that obeys to principles. There certainly remain pieces in the structure which the linguist can point at and call a word. Words cannot integrate all kinds of projections from the functional superstructure, they cannot be split up and have parts of them moved around, they cannot integrate into other words if they already projected functional superstructure, they cannot move around if the move is not motivated, etc. In short, they are built in syntax according to principles. The essence of what Distributed Morphology hypothesises is that these are nothing but independently motivated *bona fide* syntactic principles (cf. Borer 2001).

The idea that morphology is syntax does not do away with the fact that the interaction between what would call syntactic structure and morphological structure raises interesting problems. As stated above, the concept of what is a word and what is a sentence has stayed rather stable. It therefore also remains meaningful to state that syntactic structure may seem to be incorporated in word structure. Compounds, for example, may contain syntactic structure. A framework that assumes but a single structure-generating module is specifically fit to capture such data. The present work contains various examples of empirical domains that shows integrations or oppositions of word structure and sentential structure. Strictly speaking, I cannot call the morphology-syntax interaction an interface problem, given that I hypothesise they belong to the same module, but that does not take away from the fact that word structure versus sentence structure will be a recurrent theme in the work.

#### 1.4 The vocabulary-morphosyntax interface

As has been noted above, Distributed Morphology is a framework that endorses the root hypothesis, according to which the root node does not contribute or project any morphosyntactic features in the syntactic structure. The root-based nature of Distributed Morphology is perhaps the most unconventional aspect of the theory. It goes against a long-standing tradition of how a lexical item is defined in linguistics. Both psycholinguistic approaches and theoretical approaches understand lexical items as symbols that combine meaning, form and morphosyntactic features in the lexicon, as is illustrated by the following quotes from Levelt (psycholinguistics) and Chomsky (theoretical linguistics) (see also Chomsky 1965):

“The message contains one or more concepts for which we have words in our lexicon, and these words have to be retrieved. They have syntactic properties, such as being a noun or a transitive verb, which we use in planning the sentence, that is in ‘grammatical encoding’. These syntactic properties taken together, we call the word’s ‘lemma’. Words also have

morphological and phonological properties that we use in preparing their syllabification and prosody, that is in ‘phonological encoding.’” (citation taken from Levelt 1999: 223)

“The heads of these categories [VP, AP, PP, ..., M. De Belder] have (1) categorial features; (2) grammatical features such as  $\Phi$ -features and others checked in the course of the derivations, [...]; (3) a phonological matrix, further articulated by the mapping to PF; (4) inherent semantic and syntactic features that determine s(emantic)-selection and c(ategorial)-selection, respectively.” (citation taken from Chomsky 1995:54)

The idea reflects the original early 20<sup>th</sup> century proposal of conceiving the symbol as a straightforward, but arbitrary connection between a concept and a form (de Saussure 1972, first edition 1916), with the nuance that in more recent times it has been argued that every aspect of the lexical item is stored in quite intricate ways: meanings are decomposed into features and qualia (Pustejovsky 1995), form is decomposed into syllables, elements of the syllabic structure, phonemes and phonemic features are now seen as part of the symbol as well, just as syntactic features which are stored as selectional grids, theta grids, etc.

Under the assumption that open-class lexical items are marked for syntactic features, the lexicon is dominant in the lexicon-syntactic interface: words come as pre-defined syntactic packages and dictate parts of the syntactic structure. For example, a verb will come with a stored *Aktionsart*, theta roles, number of arguments, etc. In contrast, as soon as one no longer assumes any categorial or syntactic features on an open-class lexical item, syntax dominates. The syntactic structure defines how the open-class lexical item functions in the structure. The consequences of this approach have been discussed in depth in Borer’s trilogy (2005a,b,2013a), the most important contribution to roots, and in several bundles (e.g. Alexiadou, Borer and Schäfer 2014 and Doron 2014), bundles to which I have contributed as well. My doctoral dissertation (De Belder 2011a) also addressed specifically such consequences, mainly for Dutch nominal structure and derivational processes, and I refrain from going in depth into the issue here. Readers who are interested in how a morphosyntactic derivation proceeds under the root hypothesis can consult these sources.

The present work does contribute to understanding of the vocabulary-morphosyntax interface, however. More specifically, it addresses three further issues. Firstly, it addresses the theoretical question whether roots always need to merge with functional heads. It is generally assumed that they do, as they are categoriless in isolation and therefore uninterpretable. For example, without functional superstructure a by hypothesis categoriless root such as *work* would not be categorised as a verb or a noun and it is therefore unknown whether it should be interpreted with its verbal or nominal meaning. This theoretical concern goes back to Chomsky (1970:13) who points out that ‘simplifying the base (i.e. the lexical core of the derivation, M. De Belder) comes at the cost of greater complexity of transformations, or conversely’. In other words, categorial and other syntactic information needs to come from somewhere and if not from the lexicon, then from structure (see also Embick 2020 for discussion and chapter 7 in the present work on the morphosyntax-semantics interface). Despite the legitimacy of this theoretical concern, I present empirical evidence that roots can occur without functional superstructure when they occur as non-heads in Dutch compounds in chapter 2.

Secondly, there is the issue of grammaticalisation. Some vocabulary items may occur both as roots and as functional vocabulary items, a fact which deserves our attention in a framework which assumes a bifurcation in Vocabulary between root vocabulary items and functional vocabulary

items. In De Belder (2011a) I raised the issue on the basis of examples from Dutch nominal structure and the issue has been further developed in Klockmann (2017) and Cavigani-Pots (2020). I address a data set from the verbal domain in this dissertation in chapter 6. This chapter also addresses the further issue which involves the difficulty to set apart malleability from grammaticalisation. Malleability refers to conversion: it is the phenomenon that a root can surface in structures of different categories or subcategories. For example, *kiss* can be used as a verb or a noun, *chicken* can be used as a count noun or a mass noun, ... As I have discussed above, the idea is that the functional superstructure defines these (sub)categories and the root, by virtue of its categorial flexibility, is described as *malleable*. Empirically, it then becomes a puzzle to distinguish malleability from true grammaticalisation: how to tell the difference between a root that is simply showing extreme categorial flexibility versus one that truly grammaticalised into a functional vocabulary item? The present work addresses the issue by studying position and motion verbs which can surface both as agentive, lexical verbs and with bleached meanings. I argue that a subset of these verbs simply show malleability and others grammaticalised into copulas. The grammaticalisation is directly relevant for the relation between the syntax and the lexicon, malleability has consequences for the interface with semantics as well (see section 5).

Thirdly, the dissertation discusses the fact that Dutch shows a relation between syllable structure and category: Dutch nouns contain more syllables and more complex syllables than Dutch verbs. At first sight, this observation seems to falsify the root hypothesis: whether a vocabulary item is a noun or a verb seems already to be contained in its phonological structure. However, well-known facts warn against hasty conclusions. Consider Hebrew in this respect. As a Semitic language, it is commonly seen as a root-based language *par excellence*. At the same time, it is also known that Hebrew shows an important phonological noun-verb opposition: verbs need to be templatic, whereas nouns do not (Bat-El 1994). The same holds for Arabic: it is clearly root-based, yet it is at the centre of the discussion of phonological noun-verb oppositions (McCarthy 2005; Cable 2004). I argue that, indeed, if one looks closer, one can observe that the relation between syllable structures and categories is not defined at the lexicon-syntax interface, but at the syntax-phonology interface.

In short, the relation between the lexicon and syntax remains an important theme in my work as the root hypothesis is one of the most debated aspects in Distributed Morphology. More generally, Vocabulary is unavoidable in a discussion of morphosyntax and its interfaces. The word combines phonological form and meaning and it will realise syntactic nodes. It is therefore an acquired piece of information that intimately interfaces with syntax, semantics and phonology. This mere state of affairs is bound to raise interface questions.

## 1.5 The morphosyntax-semantics interface

How one defines the lexical core of a syntactic structure has not only considerable consequences for how one constructs the lexicon-syntax interface. It also has considerable effects on the syntax-semantics interface. The issue goes beyond the mere root or no roots debate. Even if one endorses the root hypothesis, the precise definition of the root's ontology, will determine how the conceptual interface should be modelled. I discuss these issues in the present section.

In the seminal papers of Distributed Morphology, the root is taken to be fully categorially and featurally underspecified. It has long been stipulated that it merely contains a feature [Root]. In more contemporary work, this feature has been given up on the favour of an index. As was already

discussed above in section 2, this abstract index refers to a vocabulary item (Harley 2014), as in (12) (taken from example (15) in Harley 2014).

(12) PF instructions:

$\sqrt{77} \leftrightarrow / \Theta row /$

LF instructions:

$\sqrt{77} \leftrightarrow$  “vomit” / [ v [ [ \_\_ ]<sub>v</sub> [ up ]<sub>P</sub> ] ]<sub>vP</sub>  
 $\leftrightarrow$  “a light blanket” / [ n [ [ \_\_ ]<sub>v</sub> ] ]  
{... other meanings in other contexts...}  
 $\leftrightarrow$  “throw” / elsewhere

The index, e.g.  $\sqrt{77}$ , then merges at the root position. For this index phonological instructions are stored at vocabulary insertion, which guarantees that the appropriate exponent will realise the root. At LF, the index corresponds to the root’s basic meaning and stored lexical knowledge such as information on collocations, idioms, etc. Importantly, the proposal does not assume a direct Saussurian connection between a phonological form and a semantic representation, as sketched in (13) (see also section 4 above):

(13) Not assumed in Harley (2014):

$/ \Theta row / \leftrightarrow$  “vomit” / [ v [ [ \_\_ ]<sub>v</sub> [ up ]<sub>P</sub> ] ]<sub>vP</sub>  
 $\leftrightarrow$  “a light blanket” / [ n [ [ \_\_ ]<sub>v</sub> ] ]  
{... other meanings in other contexts...}  
 $\leftrightarrow$  “throw” / elsewhere

The connection between form and meaning of the root is established indirectly, by referring to the same root index, contained in syntax. Given that syntax contains the index, the identity of the root can be guaranteed at the interfaces that take syntax’ output as their input. A direct connection between PF and LF, as was proposed in Harley & Noyer (1999) and as depicted in (1) (see section 2 above), is not necessary.

De Belder & Van Craenenbroeck (2015) hold a very different view on the root node. They define the root node as an empty set as a consequence of the fact that the root is the first node in the tree. When starting the derivation of a syntactic structure, the operation Select will select a first functional feature  $\alpha$  from the Numeration. This feature is included into an empty derivation. This empty object is indeed empty, i.e. it is the empty set  $\emptyset$ . The very first Merge of the structure therefore creates the object  $\{\alpha, \emptyset\}$ . This empty set then functions as the root node. In their proposal the root is thus the most unspecified node imaginable. As such, they derive that roots have no grammatical features or a syntactic category, as argued for at length in Borer (2005a,b).

An important consequence of their proposal is that the root is defined structurally rather than lexically: a root is defined as a certain position in the tree. It is a type of *syntactic node*. It is not defined by the lexical item that merges in the tree. They show that when functional items realise a root position, the functional properties of the lexical item are inert. After all, the properties of the root are defined by the absence of features in the tree, not by the vocabulary item that happens to

realise it post-syntactically. The root node thus does not contain a root feature or a root index or anything that would define the node as a root *lexically*. Their model also does not need such a feature or an index: the root node does not depend on it to be defined as a root.

In their proposal, the root node will be realised by a vocabulary item post-syntactically. This state-of-affairs has an important consequence for the way the structure interfaces with the meaning-side of the model: no index is shipped to the conceptual interface at spell-out. The conceptual interface then depends on information from Vocabulary Insertion to determine which lexeme should be interpreted at the root node. In fact, this state of affairs is reminiscent of how the model was originally designed. After all, the original root feature also did not refer to a specific lexeme and the conceptual interface thus depended on Vocabulary Insertion, an operation at PF, to receive input. This is made explicit in the original model depicted in (1) above: there is direct line between the conceptual interface and PF.

One model is not *a priori* more elegant or minimal than the other one. Harley (2014) can give up on a direct interface relation between PF and the conceptual interface, but it comes at the cost of assuming a root index in syntax. De Belder & Van Craenenbroeck (2015) need to assume the interface relation, but they postulate the most minimal root node imaginable. The issue should be settled empirically. One empirical prediction is that if syntax ships information on the lexical node to LF, the lexical core should be built completely. For morphologically complex words, this entails that they all should be derived syntactically. Yet, in chapter 2 I argue that root primary compounds in Dutch are built by the operation Fission: the operation splits the root node in more than one root node and, as a result, several root vocabulary items can be inserted. The whole surfaces as a specific type of compound. Fission is a post-syntactic operation at PF's submodule 'Morphology'. More specifically, I argue that if one accepts fission as a possibility for functional nodes, there is no reason not to assume it for the root node as well. I argue that this is indeed an empirically attested possibility, resulting in a specific type of compounding in Germanic. If one accepts these arguments, the conclusion that there must be a direct connection between Vocabulary Insertion and the semantic interface becomes unavoidable.

I have to admit that the present dissertation contains two different analyses for root primary compounds. In chapter 2 I propose they are derived post-syntactically, chapter 5 proposes a syntactic derivation. I still endorse the post-syntactic derivation. I decided not to alter this incongruity in chapter 5 because the chapter is co-authored and it has been published as printed here. As I pointed out above, whether one favours one proposal over the other strongly depends on one's view on what a root is. It will become clear that a syntactic derivation of root primary compounds entails that roots can merge with one another and project into a phrase (see Van Craenenbroeck 2014 for discussion). If one assumes that a root projects, it must contain a feature to project. Alternatively, one can propose that indices can project as well and adopt this possibility in one's theory on merge and labelling. Empty roots cannot project: a node that does not contain a feature or an index contains no element that could project (De Belder & Van Craenenbroeck 2015, Van Craenenbroeck 2014).

Recall that I discussed above that Chomsky (1970:13) pointed out that 'simplifying the base (i.e. the lexical core of the derivation, M. De Belder) comes at the cost of greater complexity of transformations, or conversely'. Indeed, the underspecified root, albeit a root feature, a root index or an empty root, is possible in Distributed Morphology because of the strict compositionality of the functional heads in the syntactic structure. If the surrounding structure is sufficiently specific, a free riding node does not undermine the interpretability and well-formedness of the syntactic

structure at the interfaces. For example, if the functional structure of a verb is fully equipped to derive a causative verb, there is no need to encode causativity on the lexical node. In fact, there is even no need to encode verb-hood on the lexical node. The avoidance of encoding information more than once in the syntactic tree is a typical tenet of work in DM.

This idea is fully adopted in the present dissertation (as it is in all my other work). A fleshed out example is presented in chapter 6. It presents, amongst other things, the distinction between the bleached locational meaning of the posture verbs *liggen* ‘to lie’, *staan* ‘to stand’ and *hangen* ‘to hang’ and their lexical versions. Hence, the verb allows for two different interpretations, the lexical reading allows the verb to be contrasted with a different posture verb, indicating that the original semantics of the posture verb is present, the bleached reading does not allow for this contrast as it lacks the original semantics of a posture verb:

- (14) A: **Staat** het danseresje van Degas in de VS? (bleached reading)  
 stands the little.dancing girl of Degas in the US  
 ‘Is Degas’ ballerina in the US?’

- B: # Nee, ze **ligt** in de VS.  
 no, she lies in the US  
 ‘No, she is lying down in the US.’

- (15) A: **Staat** de ballerina op het podium? (lexical activity reading)  
 stands the ballerina on the stage  
 ‘Is the ballerina standing on the stage?’
- B: Nee, ze **ligt** op het podium.  
 no, she lies on the stage  
 ‘No, she is lying down on the stage.’

To capture such ambiguous semantics, root-based proposals thrive. They allow one to postulate a fully underspecified lexical core, the semantic distinction is subsequently derived from a different set of functional heads. For these specific verbs, I argue that the lexical reading comes from a set of functional heads that define the root as an agentive activity, the bleached reading is a state (in the sense of Dowty 1979:184). These posture verbs are thus not *ambiguous*, they are rather semantically *underspecified*, allowing them to be malleable semantically.

If ‘simplifying the base (i.e. the lexical core of the derivation, M. De Belder) comes at the cost of greater complexity of transformations, or conversely’ (Chomsky 1970:13), the reverse holds as well: if one wants to simplify the complexity of the syntactic structure, the base will have to carry the burden. Wood and Marantz (2017) use the logic so familiar in Distributed Morphology that a certain node can be underspecified if the context is specified enough to derive meaning to reverse the proposal: they propose an underspecified transitivising head, which inherits a category form the first constituent it merges with. This head can merge with all kinds of categories, it is categorially chameleonic: it derives its category from its context. In technical terms this is called contextual allosemy. For example, in the context of a locative PP it may be interpreted as a head that introduces a figure, in the context of an agentive vP it may be interpreted as a head introducing an agent. In short, the only thing the head does, in essence, is introducing an argument. How that argument can

be most meaningfully interpreted is determined by the syntactic features in its context. This context is not restricted to purely syntactic features, it may also depend on roots. The underspecified transitivising head may merge, for example, with lexical roots to derive prepositions (some prepositions are seen as roots in this proposal) and High Applicatives. As such, even lexical information may affect the interpretation of the transitivising head. Its semantics is thus specified by the surrounding projections, including the root. Importantly, the meaning partially depends on stored possibilities at Encyclopedia and less on compositional features. Meaning is derived from the syntactic and lexical context and not necessarily from a strict composition of features ('contextual allosemy', Wood 2015).

The debate whether the root should be underspecified or whether functional heads, such as transitivising heads, should be underspecified is discussed in depth in chapter 7. In the remainder of the chapters, it becomes clear that I am not Switzerland in this debate. My work has always clearly endorsed a minimal root and a full compositional, monotonously built semantics derived from the functional structure. The present dissertation does not leave this path. A clear example is the proposal in chapter 6, which distinguishes between the two readings in (14)-(15) simply by assuming different functional heads above the root. Further arguments for the minimal root are presented in chapter 2.

In short, I endorse a minimal view on the root and fully compositional functional structures that define the semantics of the structure. Given that I endorse the view that the precise lexemes are only selected post-syntactically, I assume a direct interface between PF and the conceptual interface as well.

## 1.6 The morphyntax-phonology interface

In Distributed Morphology there is little consensus on how morphosyntax and phonology interface, the reason being twofold. The first reason is a theoretical concern: some central tenets of Distributed Morphology are hard to reconcile with the principles of Optimality Theory, the contemporary dominant approach to phonology in generative linguistics. Below I sketch the problem in some more detail. The second reason is the fact that the empirical domains at this interface show a lot of internal variation, which has led to the debate for which module or interface they are relevant in the first place. Needless to say, different answers to this basic question have resulted in very different accounts. For example, if one concludes that allomorphy is a problem for Vocabulary one proposes a very different account than if one concludes it is a phonological problem requiring a phonological derivation. The present work does contribute a great deal to this second issue. I discuss three empirical domains from Dutch and I offer solutions and suggestions as to where and how they should be analysed. These domains are allomorphy, epenthetic consonants and category-specific syllable structures, which were already mentioned in section 4. Before addressing these empirical domains, I will first briefly discuss the theoretical tension between Distributed Morphology and Optimality Theory.

Theoretically, there is a tension between the piece-based, realisational nature of Distributed Morphology and the field's dominant approach to phonology, i.e. Optimality Theory, which evaluates phonological words as a whole and compares competing candidates. The problem is discussed at length in Embick (2010), the following quote summarises the problem:

### The schism

Globalist theories of morphophonology [i.e. approaches such as Optimality Theory, M. De Belder] require competition between multiple potential expressions of a given input. According to the localist morphosyntactic theory, this is impossible because the competitors are not derived by the grammar. (Embick 2010:11)

Indeed, Optimality Theory's operation *Eval* evaluates various candidates and it is unclear in Distributed Morphology which operation would generate these candidates from the output of Vocabulary Insertion. Yet, the problem does not seem unsolvable to me. One should essentially guarantee that the output of vocabulary insertion is a phonological representation that is abstract enough such that various surface representations could be generated from it. The problem could then be remedied by integrating an appropriate mechanism in the model (see, for example, Bye and Svenonius 2012 for a proposal). In the same vein, in chapter 8 I propose that the output of Vocabulary Insertion can be a disjunctive set of candidates ("insert A or B"), which can then serve as input candidates for Phonology (see there for discussion and a comparison with Bye and Svenonius 2012). Admittedly, such a disjunctive set still differs from a full-blown candidate set as is needed for a framework such as Optimality, but it is a step in the direction of a compromise.

A more difficult theoretical tension between Distributed Morphology and Optimality Theory described by Embick (2010: chapter 6) is the fact that allomorphic choices are determined locally. He describes the hypothetical situation sketched in (16) for which there would be two phonologically determined allomorphs available to realise X (Embick 2010:23):

(16) Root-X-Y-Z

A local theory would predict that the allomorph of X is determined on the basis of what is available at PF. Under the assumption that the root and X are a phase and spelled out together, the phonology of the root will determine the selection of the appropriate allomorph for X, even when this selection will appear to be less optimal choice once Y and Z are merged and the phonological word is completed. (This could happen, for example, when the phonological word's foot structure has consequences for the most optimal phonological form of X). A global theory, in contrast, would evaluate the phonological word as a whole and would thus take into account the phonological shape of the entire word, rather than strictly local information. Embick (2010) argues that the local approach is the approach that makes the correct empirical predictions.

In sum, the two approaches may make very different empirical predictions because of a very different view on which pieces of the phonological word are taken into account when phonology evaluates a form. It has proven difficult to find a solution to this tension without far-reaching changes in either the assumptions on morphosyntax or phonology. Some solutions have proposed to interweave morphological and phonological requirements. Morphological selection mechanisms are then redefined as constraints within Optimality Theory (Prince & Smolensky 1993,1997). Other solutions involve redefining Optimality Theory's operation *Eval* as an evaluation that operates on incremental morphosyntactic steps of the structure rather than on the phonological word as a whole (e.g. Wolf 2008). Future will tell which direction the field will settle on.

Allomorphy is a surprisingly hard domain to tackle. As discussed above, it is a central piece of data that illustrates the fundamental schism between Distributed Morphology and Optimality Theory. The problem does not stop there, however. It is also a central empirical domain in the

debate which module or interface should tackle morphophonological irregularities, which I called the second big problem when it comes to understanding the morphosyntax-phonology interface.<sup>7</sup> Morphophonological irregularities involve the fact that the straightforward match between morphosyntactic structure and vocabulary items may show disturbances at the surface. Allomorphy, for example, instantiates a case where several exponents would have matched a single morphosyntactic node. Syncretism instantiates a case where a single exponent matches various types of morphosyntactic nodes. Epenthetic consonants, for example, illustrate a case in which it is *a priori* not even clear whether the exponent realises a morphosyntactic node.

Such cases present a challenge for the interface between morphosyntax and the various operations at PF. The problem at hand is that it is often not even clear which module a surface irregularity should be ascribed to. What caused an allomorph or an epenthetic consonant to appear? Was the morphosyntactic structure marked in a certain way such that a perhaps at first sight less expected exponent was selected, was there perhaps something irregular in the way vocabulary items were organised at vocabulary or did the unexpected exponent simply match the phonology better?

The present work discusses three empirical domains which raise important issues on the interaction between vocabulary, morphosyntax and phonology. The first one is the infamous domain of allomorphy, infamous indeed as it is well-known that allomorphy is often phonologically optimising, but certainly not always. The second one is the epenthetic consonant: it is typically phonologically optimising, but it may be determined by lexical and syntactic factors as well. The third one is the observation that syllable structure in Dutch is category-specific. For all three domains it is *a priori* not clear which modules are involved in the selection of the exponents. I discuss them in more detail below.

Consider first allomorphy. As pointed out by Bonet and Harbour (2012: 220), there is little agreement in the literature as to which mechanism is responsible for the selection of an allomorph. Indeed, proposals range from stipulations or rankings at Vocabulary Insertion to morphological and phonological constraints in OT-based models. Bonet and Harbour point out that the source of the theoretical dispute probably stems from the disparity in the data (see also Bye 2007). Indeed, the occurrence of a specific allomorph may be determined by its lexical context, its syntactic context, or its phonological context. An example of lexical selection is given in (17): the participle of the verb *uitverkiezen* ‘to select’, as shown in (17)a may occur in an exceptional form in a specific lexical collocation, as in (17)b:

- |                               |  |
|-------------------------------|--|
| (17) a. Jij bent uitverkozen. | b. het uitverkoren volk                      |
| you are selected              | the selected people                          |
| ‘You are selected.’           | ‘the Chosen People (i.e. the Jewish people)’ |

An example of syntactic selection is shown in (18): the Yiddish allomorph *kinder* ‘child’ is limited to syntactic plural contexts, in other contexts the default allomorph *kind* ‘child’ will appear (Lowenstamm 2007). To understand the reasoning example, one should note that the actual plural affix in (18)b is assumed to be a zero affix and not the string *-er*. This assumption is based on the analogue with (18)d, which contains an overt plural affix (*-ex*) and which shows that *kinder-* occurs inside the diminutive. Diminutive affixes merge outside plural affixes (De Belder 2011b, De Belder,

---

<sup>7</sup> Allomorphy is also the empirical domain that caused a debate on whether the root’s index is an abstract index or a phonological index (Harley 2014 versus Borer 2013).

Faust and Lampitelli 2014), hence *kinder* cannot contain the plural affix and is analysed as an allomorph.

(18) a. kind	b. kinder-Ø	c. kind-l	d. kinder-l-ex
child	children	child-DIM	child-DIM-PL
'child'	'children'	'little child'	'little children'

Example (19) shows phonological selection: the choice between the two allomorphs (-en or -s) of the Dutch plural affix is *amongst other things* based on the trochaic nature of the language: the affix that results in the phonological word being a trochee will be preferred (Van Haeringen 1947).

(19) a. een kat	[kat]	b. twee kat-en	[katə(n)] <sup>8</sup>
a cat		two cat-PL	
'a cat'		'two cats'	
a. een otter	[ɔtər]	b. twee otter-s	[ɔtərs]
an otter		two otter-PL	
'an otter'		'two otters'	

This threefold empirical distinction in the selection of allomorphs has led to a longstanding discussion as how to organise the selection of allomorphs theoretically and technically. Either one attempts to unify the selection of these three different types of allomorphy in a single account, which is doomed to be less than elegant, given that it has to capture three very different types of data, or one gives up on a unified account and one distributes the selection of allomorphy over various modules. To make matters even worse, the data are often ambiguous. Consider the selection of the allomorph of the plural affix again. Even though there is a general consensus that the selection is indeed based on the trochaic foot nature, this does not completely capture the data. Firstly, the choice of the affix may also depend on the final vowel of the root if the last syllable is an open syllable: front vowel select -en (*idee-en* 'ideas'), back vowels select -s (*taboe-s* 'taboos') (Kooij & Van Oostendorp 2003: 96). Secondly, there is some regional and lexical-based variation. For example, the word *test* selects -en in Southern Dutch, as expected, but -s in Northern Dutch. Thirdly, the lexical stratum has an influence: loanwords select -s, even though there are stored exceptions (as *test* for example). Fourthly, kinship names may unexpectedly select -s as well in all variants (p.c. Gertjan Postma):

(20) a. een oom	[om]	b. twee oom-s	[oms]
an uncle		two uncle-PL	
'an uncle'		'two uncles'	

It is *a priori* unclear whether this should be ascribed to a lexical irregularity or to morphosyntax (kinship names may be associated with a specific syntactic structure). Perhaps other factors are relevant as well, but I have illustrated the point: everything seems to be at play simultaneously. The

---

<sup>8</sup>The pronunciation of the final /n/ of the plural marker is optional and depends mostly on the speaker's dialect and the phonological context.

choice of the affix depends on the lexicon, phonology, perhaps morphosyntax, ... This illustrates the persistent problem with allomorphy. Even when trying to capture something as basic as the Dutch plural affix, already at the level of observational adequacy, one runs into chaos. It is impossible to point at a single module to neatly derive all instances of allomorphy. It should not come as a surprise that this leads to considerable debate when it comes to descriptive and explanatory adequacy. In chapter 8 I discuss a specific set of data in detail and I argue that allomorphy selection is distributed over vocabulary insertion and phonology.

An empirical domain similar to allomorphy is the case of so-called syntactic epenthetic consonants. In her overview of types of epenthetic consonants, Zygis (2010) points at the existence of what she calls syntactic epenthetic consonants. These are epenthetic consonants that are motivated phonologically - as epenthetic consonants are- and only selected in a specific syntactic and/or lexical environment. For example, the epenthetic consonant [t] in (21) depends on the suffix *-lich* to appear (example taken from Zygis 2010: 112, her example (2)).

- (21) *wesen-t-lich*  
STEM-EPENTHETIC.CONSONANT-SUFFIX  
'considerably'

The immediate first question concerning such epenthetic consonants is whether the selection is based on a specific vocabulary item (e.g. the affix *-lich*) or on the morphosyntactic node (i.e. the morpheme in terms of Distributed Morphology) that gets realised by the vocabulary item or perhaps both. The second question is how the interactions between syntax, vocabulary and phonology need to be modelled as to allow for the insertion of the epenthetic consonant. Chapter 10 presents a proposal for the Dutch epenthetic consonant [s] that appears in substandard diminutives, such as *dag-s-ke* 'short day'.

A third empirical domain of concern has gotten less interest within Distributed Morphology. It involves the observation that syllable structure may be category-specific, as already described in section 3. (Note that category-specific phonology in general was given attention, see Bobaljik 1998 versus McCarthy 2005, Cable 2004). As mentioned above, I present empirical evidence for a connection between the category of a lexical item and its syllable structure in Dutch: I show that for Dutch nouns the probability is higher to have more and more complex syllables than it is for Dutch verbs. I mentioned that this observation is somewhat surprising under the root hypothesis. If roots are categoriless, then how come it would possible to guess their category on the basis of their syllable structure? I argue that the connection is not between the lexicon and morphosyntax, but between morphosyntax and phonology. In other words, roots are not classified for their category on the basis of their syllable structure in the lexicon, but vocabulary insertion of roots in certain syntactic positions can be hindered by a lower phonological optimality.

In sum, for all these domains we are facing challenges that can be summarised as follows. One needs to determine what is a phonological phenomenon and what is not and one often has to tease apart lexical selection from syntactic selection. Once the data are sorted, one has to be able to derive them technically, which is less self-evident, given the theoretical tensions between Distributed Morphology and Optimality Theory. These empirical, technical and theoretical issues are addressed in the dissertation.

## 1.7 Chapters in the dissertation

### 1.7.1 Compounding (chapters 2 to 5)

The dissertation contains three articles on primary compounding in Dutch and one article on phrasal compounding. Primary compounding is the ‘basic’ type of compounding: the non-head contains only one root and the right-hand part is the head. I argue in these articles that (at least) three subtypes of primary compounding can be distinguished in Dutch. They are illustrated in (22).

- (22) a. kreeft-woord    b. kreeft-en-soep    c. Kreeft-s-keerkring  
      lobster-word      lobster-EN-soup      lobster-s-tropic  
      ‘palindrome’      ‘lobster soup’        ‘Tropic of Cancer’

I argue that the subtype illustrated in (22)a contains a bare root as its non-head. I call this type *root primary compounding*. Empirically, the bare root results in the absence of any so-called ‘linking element’. After all, the root is bare and did not merge with anything that could subsequently be overtly realised by an affix. Furthermore, the non-head is not necessarily interpreted as nominal (even though it would be in this example). For the subtype in (22)b I argue that the non-head contains a nominal functional head, more specifically a nominal class projection. I call this type *nominal primary compounding*. The nominal functional projection is realised as either *-s-* or *-en-*, depending on the specific root in the non-head (for *kreeft* ‘lobster’ *-en-* is selected, as can be seen in the example). Needless to say, the functional head defines the non-head as nominal and it is interpreted as such. For the subtype in (22)c I argue that the non-head contains a determiner, which surfaces as *-s-*. This results in the non-head having referential properties and I call the type *referential primary compounding*. Again, this non-head is interpreted as inherently nominal due to the nominal functional projection. The proposal that the non-head can be referential is quite a novel claim: referential non-heads were assumed to be principally excluded from compounds in any language. More generally, the identification of these three subtypes for Dutch primary compounding and the analyses I provide are novel approaches to Dutch compounding, even though I am, admittedly, certainly not the first one to suspect that (22)a and (22)b should be set apart (see Bilderdijk 1826:207 who already formulated a brief remark).

Chapter 2 on root primary compounding involves the morphosyntax-lexicon interface with consequences for semantics. The non-head of this type of compounding is fully acategorial. It does not contain categorial heads (i.e. little heads) or functional projections. Interpretationally, the category of the non-head remains unresolved syntactically (and depends on world knowledge to be interpreted). In some cases, this may actually result in ambiguity:

- (23) antwoord-centrum  
      answer-center  
      ‘center where answers are stored / center where answers are provided/...’ (nominal readings)  
      ‘center where one answers the phone/...’ (verbal reading)

The discussion supports the hypothesis that the most minimal lexical projection is a root (Halle and Marantz 1993, Borer 2005a) and that it occupies a syntactic slot in the structure which does not need to be licensed by superstructure in order to be semantically interpretable or realizable. This is an important finding, as it was assumed that bare roots are principally excluded, as they would not be

interpretable (e.g. Harley 2012, Starke 2009 i.a.). These compounds falsify this assumption and demonstrate that categorial assignment is not a strict requirement for interpretation. Admittedly, the end result may be ambiguous, but in its ambiguity, it is perfectly interpretable: the structure simply ends up having several interpretations rather than none.

The article also contains considerations on how to recognize morphological complexity in the first place. In my opinion, this problem is often overlooked. In order to achieve a correct understanding of the vocabulary-morphosyntax interface, it is important to recognize the relevant data. Needless to say, morphologically simplex words are not relevant to such a study.

The article also argues for a direct relation between PF and LF. It will become clear that the morphosyntactic derivation is not assumed to contain a root index, but an underspecified root and that roots are only inserted at vocabulary insertion. Such an analysis necessitates a direct connection between PF and the conceptual interface, otherwise the conceptual interface would not have access to the contents of the roots.

**Chapter 3** discusses so-called linking elements. Cross-linguistically, they may actually fulfil two very different roles: they may be phonotactic elements or they may realize a morphosyntactic head. As such, they are in instance of data for which it is not *a priori* clear for which module or interface they are relevant in the first place. I present a case study of linking elements in Dutch which do fulfil a morphosyntactic role. I note first that when ‘linking elements’ are of this type, the linguist should be aware of the fact that they may be null morphemes. I show that, indeed, Dutch may have such a null affix in its inventory. I then argue that the Dutch linking elements are in fact noun class markers and as such, I derive the fact that two of them are identical to plural marking through cumulative exponence. I point out that these noun classes co-exist with gender marking and may cause so-called erroneous reference.

The analysis shows that a careful study of such elements provides us with new insights into two domains simultaneously. Firstly, we learn more about the morphosyntax of the compound’s left-hand part, the structure of the compound itself and the inventory of compounding types in the language. The importance of the left-hand part cannot be underestimated in my opinion. It is a unique domain: it shows how a category behaves functionally when it does not function as an argument. As such, we learn, secondly, something about the morphosyntax of a category. In this contribution it appears that the Dutch nominal domain has nominal classes which are less visible in the nominal group as they are syncretic with number marking. Take away number marking and they become visible.

Given that noun class marking is a syntactic projection, it is argued that a syntactic projection is contained within the compound. Under the assumption that morphology is syntax, such a state of affairs is expected. It becomes clear that there seems to be a selectional relation between a root and a noun class, just like there may be a selectional relation between a root and a gender class. The proposal is that the connection is stored at Encyclopedia. Syntax may generate just anything, but not everything is interpretable at the end of the day. Only syntactic structures that match stored interpretations (e.g. listed combinations of a root and its gender or its class) are interpretable.

**Chapter 4** is another study on nominal projections within the compound. It is generally assumed that non-heads of compounds cannot be referential: they are assumed to lack a D-layer. In this article I argue that this assumption is false. Dutch has compounds with a referential non-head. These non-heads belong to the types of lexemes Longobardi (1994) identified as typically subject to N-to-D raising: proper names, kinship names, names of days of weeks and Bank holidays, unique entities and contextually unique entities. I argue that the compounds contain a nominalizing head

specified for uniqueness and referentiality, which I call little d. The non-head of the compound incorporates into little d. I compare the morphosyntax of these non-heads to the morphosyntax of modificational compounds, of proper names in argument position, of the Dutch -s possessive and title expressions, which I argue are Dutch referential construct states.

Together with chapter 2 and chapter 3 the work studies the kind of projections that are allowed on the non-head of the compound. This domain is rather unexplored and I would like to draw attention that several nominal heads can be contained in this domain and that their presence or absence goes hand in hand with relevant semantic distinctions. The semantics of compounds is thus not simply listed: it shows predictable compositional semantic patterns. This is reminiscent of McGinnis' (2002) work on idioms: semantic meaning may be stored, but structural meaning cannot be overridden. A bare root is categorially ambiguous, a nominal class projection results in a modificational compound and a referential layer results in a referential compound,

Chapter 5 is the final chapter on compounding. The three articles on primary compounding are followed by a contribution on ANN-compounding in Dutch, i.e. compounds which have an adjective plus noun as their non-head and a nominal head. In this contribution Marjo Van Koppen and I provide empirical evidence against the claim that morphology contrasts with syntax in dealing with items that are listed in the lexicon. (Jackendoff 1975, Aronoff 1976, Jensen and Stong- Jensen 1984). More specifically, we distinguish between three types of ANN compounds in Dutch. We show that the structural properties of these types do not show a one-to-one mapping with lexical properties, such as having a listed or even idiomatic meaning (see DiSciullo & Williams 1987). On the basis of this, we argue that conclusions on the structure of certain morphologically complex word-forms should be based on structural properties and not on lexical properties such as idiomacity or being lexicalized. We propose a syntactic derivation for all types of ANN compounds in Dutch (*pace* Ackema and Neeleman 2004). Structural differences follow from the level of merge: what we traditionally call morphology is syntax below the functional domain.

### **1.7.2 Posture verbs, motion verbs and copulas(Chapter 6 )**

It has been observed in typological literature that posture verbs often grammaticalise into copulas (Newman 2002). Article 6 presents an account for this phenomenon by discussing the phenomenon in Dutch and other continental West-Germanic varieties, viz. High German, Low German and Frisian. More specifically, it contrasts malleability with grammaticalisation in the empirical domain of posture verbs and motion verbs. These verbs can all be used as lexical verbs. However, they also show bleached uses, which look like copulas. For some verbs, this similarity with copulas merely stems from the fact that they can occur in non-agentive, stative contexts, which contrasts with the agentive activity reading in their lexical use. I conclude that these aspectual features thus do not inherently belong to the verb, but show malleability. Yet, other verbs indeed have grammaticalised into copulas. The article thus studies the interface between syntax and vocabulary: it distinguishes between polysemy through grammaticalization and storage of a functional vocabulary item in vocabulary and mere malleability of vocabulary items. It also contributes to the view according to which semantic is built compositionally through features in the morphosyntactic structures (see also 7 for discussion).

### **1.7.3 The syntax-semantics interface in Distributed Morphology (Chapter 7 )**

Chapter 7 gives an overview on how the syntax-semantics interface is understood in Distributed Morphology. Literature in Distributed Morphology tends to adhere to the principle of full compositionality at the LF-side: only a few authors have proposed post-syntactic structure-altering operations at the semantic side of the computation. Yet, the field does not generally assume a simple match between syntactic and sematnic structure due to underspecification of the syntactic structure. The idea that syntax can be semantically underspecified has been explored in various ways. The idea goes back to the root node, which has been seen as underspecified since the seminal papers. This idea gradually spread to little v and, more generally, transitivising heads. As suchm some authors moved away from the original idea that there is a close, precise and deterministic match between the morphosyntactic features of the structure and its meaning.

### **1.7.4 Allomorphy (Chapters 8 and 9 )**

Article 8 argues that the selection of allomorphs is distributed over two modules, viz. Vocabulary Insertion and Phonology. This is done on the basis of a case study of vowel length alternating allomorphs in Dutch. The data show a split pattern: some empirical domains can be fully captured by phonological principles. For these cases, the phonologically most optimal allomorph will be selected. In other empirical domains, phonological principles still account for many of the attested data. Yet, one attests lexicalised exceptions as well, which are clearly phonologically non-optimal. The data echo opposing views in the literature: some proposals attempt to reduce allomorph selection to phonology, others focus on the fact that many examples are simply not phonologically optimal and suggest that allomorph selection should not be done by Phonology. I argue that the opposing nature of these two types of data is actually indicative of the way the selection of allomorphs is organised. More specifically, both Vocabulary Insertion and Phonology can determine the selection of allomorphs. Vocabulary Insertion is responsible for stored information, Phonology is responsible for phonologically optimising patterns.

Article 9 shows that Breton plurality provides further empirical support for the hypothesis that the appropriate locality condition for allomorphy is head incorporation. Given that head incorporation defines word-hood, there is a close link between word-hood and the appropriate domain for allomorphy. It has been commonly assumed that the trigger of allomorphy should be sufficiently local to the morpheme subject to the allomorphy (Embick 2010). Bobaljik (2012:68) adopts as a working hypothesis that this locality should be defined as head incorporation: a head that conditions root allomorphy must be in the same complex head (i.e. the same morphological word) as the root. As such, he captures the fact that comparative suppletion is blocked in periphrastic comparatives. It is this assumption that receives empirical support in chapter 9.

### **1.7.5 Epenthetic consonants and syllables and categories (Chapters 10 and 11 )**

The bundle closes with two articles that show that syntactic features may determine phonological constraints: syntactic structure may determine the insertion of an epenthetic consonant or the acceptability of a syllable structure.

Chapter 10 studies the insertion of an epenthetic consonant determined by the affix *-ke*. The substandard Dutch diminutive suffix *-ke* occurs in words of four different categories: nouns, predicatively used adjectives, adverbs and interjections. In nouns, adjectives and adverbs the suffix may trigger an epenthetic consonant, which is not available for the interjections. This seems to go

hand in hand with a syntactic distinction: for the interjections the suffix is not to be analysed as a diminutive, but as an honorific. I argue that in all contexts, the suffix realises scalar features, but of different natures. The epenthetic consonant depends on locality with the scalar feature.

Chapter 11 questions whether Trommelen's observation on a relation between Dutch syllable structure and syntactic categories can be confirmed by means of contemporary experimental tools. Trommelen (1989) observed that Dutch nouns and verbs seem to differ in syllable structure, but this hypothesis has never been fleshed out, nor observationally or explanatorily. If Trommelen was right, Dutch would be one of the languages which show suprasegmental noun-verb dissimilarities (Smith 2011). The article indeed presents evidence for phonological noun-verb dissimilarities in Dutch (see also Trommelen 1989; Don and Erkelens 2006). We were able to confirm that verbal stems contain less syllables and less complex syllables than nominal stems through corpus-based computational research. In a large-scale online experiment, we also found that native speakers have intuitions on the word class of a pseudo-word on the basis of its syllable structure. Again, the intuitions are steered by the syllables' complexity and the number of syllables. Importantly, the experimental results show somewhat different patterns than the corpus research. This indicates that something steers the intuition in the linguistic competence that transcends mere awareness of statistical patterns in the lexicon. These results are discussed in the light of the Optimal Paradigms approach, which proposes an explanation for similar facts in other languages by pointing at inflection differences between verbs and nouns (McCarthy 2005; Cable 2004). Based on our findings, inflectional differences cannot explain the Dutch facts and we falsified the hypothesis that the paradigm as a whole can put pressure on the form of the stem in an experiment. As such, the study supports Bobaljik's (2006) scepticism on referring to paradigms when accounting for phonological noun-verb dissimilarities. We also German conversion data that show that the relevant interface relation to understand this domain is not to be situated between vocabulary and syntax, but between syntax and phonology.

## 2 The Root and Nothing but the Root

### Root primary compounds in Dutch

#### Abstract

This article is an addendum to recent contributions on the structure of compounds in root-based frameworks (Borer 2011, 2013:chapter6, to appear, Harley 2009). It presents a subtype of Dutch primary compounds of which the non-head is demonstrably a bare root. The non-head of this type of compounding is fully acategorial. It does not contain categorial heads (i.e. little heads) and neither is it categorized otherwise. As such, the discussion substantiates the root-hypothesis (Halle and Marantz 1993, Borer 2005a,b, 2013a) and it supports the view that the root occupies a syntactic slot in the structure (pace Ramchand 2008, Starke 2009) which does not need to be licensed by superstructure in order to be interpretable or realizable (pace Arad 2005, Marantz 2007).

#### 2.1 Introduction

A derivation typically consists of functional structure which merged on top of a root. As a result, roots are rarely attested in the absence of functional projections. For example, the DP in (1) only enables us to observe a root which is embedded under nominal functional structure such as number marking and a determiner. It does not provide us any access to the bare root *straat* ‘street’.

- (1) de straat-en  
the street-PLURAL  
'the streets'

In this respect the non-head of a primary compound<sup>9</sup> might be a unique syntactic domain. There is no obvious reason why such a non-head should necessarily merge with any functional structure. Hypothetically, it is therefore possible that the non-head *straat* ‘street’ in (2) is a bare root.

- (2) straat-kat  
street-cat  
'street cat'

The non-head of the compound in (2) now becomes an important testing ground to determine the theoretical status of roots. If one proposes that the core of a lexical projection consists of nothing but an acategorial root (Halle and Marantz 1993, Borer 2005a,b, 2013a), one expects the existence of compounds of which the non-head consists of anything else than a root terminal node. After all, it is the most parsimonious structure imaginable for the non-head in contemporary reasoning and there is no *a priori* reason to exclude this possibility. In this article I argue that this prediction is

---

<sup>9</sup> Primary compounds are the simplest compounds. They are not synthetic compounds and their left-hand part is not phrasal. They go by the name root compounds as well.

indeed borne out. I will present compounds from Dutch, a language in which a plethora of compounding types can be found (Booij 2002, Don 2009), and I will show there is a subtype of primary compounds in Dutch of which the non-head is indeed a bare root. It will become clear that nothing else is contained in the non-head of this type of compound, not even categorial heads (i.e. little heads) or nominal functional projections.

This article should be understood as an addendum to recent contributions on the structure of compounds in root-based frameworks. Present root-based proposals are successful in deriving categorial restrictions on the compound's non-head. There are indeed data for which building in categorial restrictions seems to be a main concern. For example, the non-head of English compounds cannot belong to just any category. The possibilities are restricted to the ones given in (3) (Selkirk 1982:14), a restriction which needs to be captured in root-based accounts.

(3)	NN	NA	PV
	AN	AA	
	VN	PA	
	PN		

Clearly, one cannot simply postulate that the non-head is an acategorial root in English compounds, as this would result in overgeneration. The non-head root is categorized. Proponents of the root therefore have formulated proposals in this direction. Harley (2009) argues for a categorial head, i.e. a little head such as  $n^\circ$ ,  $v^\circ$  or  $a^\circ$ , above the non-head root.<sup>10</sup> Borer (2013a,b) proposes a Compound Frame, i.e. a specific structure which renders the non-head equivalent to a noun.<sup>11</sup> However, if we restrict the discussion to compounds of which the non-head is categorized, the domain of compounding does not seem to benefit theoretically from the theory on roots, quite on the contrary. The compound's non-head is still assumed to be a categorized projection. In this article I therefore would like to emphasize on the fact that there exist compounds of which the theoretical status of the non-head directly supports the root-hypothesis.

The theoretical consequences of this article are not limited to evidence in favor of the root-hypothesis. Crucially, it is not only argued that the most minimal lexical projection can be a root, but also that this root can be bare. Consequently, a root can occupy a position which does not depend on categorial or functional heads to be interpretable or to be realized. This conclusion is problematic for the view which is held in Distributed Morphology according to which roots depend on a categorial head in order to be interpretable (see Arad 2005, Marantz 2007, Harley 2012). The root in the non-head of a Dutch root primary compound does not merge with a categorial head, yet the compound is interpretable.<sup>12</sup> It is equally problematic for frameworks which claim that roots merge high in the structure as a modifier of functional structure (Harley 2005, Den Dikken 2008) or that they depend on functional superstructure in order to be realized. Such a view is defended by proponents of Nanosyntax (Starke 2009, Ramchand 2008). They argue that vocabulary items of lexical categories do not spell out a root position. They rather realize a set of functional heads. The occurrence of bare roots devoid of functional structure in the non-head, i.e. a left branch, of Dutch root primary compounds is an empirical problem for such a theoretical view.

---

<sup>10</sup> She further proposes that the root incorporates into this head.

<sup>11</sup> The equivalence should be understood distributionally.

<sup>12</sup> It still holds that roots need to be embedded in a syntactic structure in order to be interpretable: the structure determines the meaning of the root at Encyclopedia.

I will proceed as follows. In the next section I first present two types of primary compounds in Dutch. I will distinguish between a first type which invariably nominalizes the root in its non-head and a second type of which the non-head is category-independent. This second type is the empirical basis of this article. In section 3 I argue against intervening categorial heads, in section 4 I present counter-evidence against the assumption that the non-head of this second compounding type is nominalized. Section 5 concludes.

## 2.2 Two types of primary compounding in Dutch

In this section I argue that there are two different types of primary compounds in Dutch<sup>13</sup>. They can be distinguished empirically and historically. It will become clear that a first type invariably contains a root which is nominalized by means of nominal functional material as its non-head. The non-head of the second type, on the other hand, does not need to be co-extensive with a noun<sup>14</sup>. This type will be analyzed in detail in the present article. I will demonstrate that its non-head is invariably a bare root.

The non-head of the first type is invariably nominal and it is followed by what is called a linking phoneme in descriptive work on Dutch morphology (e.g. Haeseryn et al. 1997, de Haas & Trommelen 1993), as is shown in (4).

- |                           |                    |                |                 |
|---------------------------|--------------------|----------------|-----------------|
| (4) a. varken-s-hok       | b. bakker-s-winkel | c. kat-en-luik | d. peer-en-boom |
| pig-LP <sup>15</sup> -pen | baker-LP-store     | cat-LP-panel   | pear-LP-tree    |
| 'pig's pen'               | 'bakery'           | 'cat door'     | 'pear tree'     |

The linking phoneme is selected by the specific root in the left-hand part and it is consistent in the sense that it can be predicted for newly formed compounds within a given dialect. This type of compounding has been attested in Dutch since the sixteenth century and it only became attested frequently since the twentieth century (Tiel, Rem and Neijt 2011:132). It has been studied extensively. As its non-head is invariably nominal, it stands to reason that the linking phoneme<sup>16</sup> instantiates a piece of nominal inflection, such as (a remnant of) case ending (Booij 2002), plural marking (Neijt and Schreuder 2009) or noun class marking (De Belder to appear). In the remainder of this article I will mainly ignore this type of compounding. I will only discuss it in contrast with the second type. When I do refer to these compounds I will call them nominal primary compounds or NPCs, as they are primary compounds with a nominal non-head.

The second type of compounding has been given less attention, but it will be the empirical core of the present article. Some examples are given below.

---

<sup>13</sup> This observation goes back to a short note in Bilderdijk 1826:207 who observed that the nominal non-head of Dutch compounds may come in two varieties, viz. with or without a linking phoneme. We will see that the present claim is more general as it includes non-heads which are not co-extensive with nouns as well.

<sup>14</sup> Most approaches to Dutch morphology sort compounds according to the (assumed) category of the non-head (see, for example, de Haas and Trommelen 1993 and Don 2009). As a result, they distinguish more than two types.

<sup>15</sup> LP = linking phoneme

<sup>16</sup> The term *linking phoneme* is thus a misnomer.

- (5) a. kleer-kast      b. speur-hond      c. snel-trein      d. achter-deur  
     cloth-closet      track-dog      fast-train      back-door  
     ‘wardrobe’      ‘tracking dog’      ‘high-speed train’      ‘backdoor’
- e. drie-luik      f. ja-woord      g. of-poort      h. niet-verjaardag  
     three-panel      yes-word      or-gate      not-birthday  
     ‘triptych’      ‘marriage vows’      ‘or-gate’      ‘unbirthday’
- i. frrrpl-geluidje  
     frrrpl-sound  
     ‘frrrpl sound’

It will become clear that the non-head of this type is invariably directly adjacent to the head, without any intervening material and that it does not need to be co-extensive with an existing noun (see sections 3 and 4 for careful discussion).

Most often, the head will be nominal as in the examples above, but this is not a requirement, see (6).

- (6) a. Zij **zweef-vlieg-t**  
     she hover-fly-INFL  
     ‘She glides.’
- b. een **kakel-bont-e**      trui  
     a cackle-colorful-INFL sweater  
     ‘a gaudy sweater’
- c. een **hoog-zwanger-e** collega  
     a high-pregnant-INFL colleague  
     ‘a late pregnant colleague’

This type of compounding is attested in the oldest Dutch texts which are available to us. This means that it is at least as old as the ninth century (Tiel, Rem and Neijt 2011, Ruissen 2011:55). Old Dutch examples of such compounds are given in (7) (Ruissen 2011:55 and *Oudnederlands woordenboek*).

- (7) a. kuo-smero      b. frīt-hof      c. hēt-muodi      d. ēn-gimi  
     cow-smear      enclose-garden      hot-mood      one-winter  
     ‘butter’      ‘atrium’      ‘anger’      ‘an animal of one year old’

Even though this type of compounding is very old, it does not differ from other compounds in Dutch in being fully productive. The examples in (8) are only added recently to Dutch vocabulary.

- (8) a. koop-moeder<sup>17</sup>  
     buy-mother  
     ‘woman who receives a child from a surrogate mother’

---

<sup>17</sup> This example was added to *van Dale* dictionary in 2009.

- b. mee-moeder  
with-mother  
'co-mother, i.e. lesbian partner of a biological mother'

On a par with other types of Dutch compounding, the meaning of the compound as a whole may be clearly related to the meaning of its subparts, as in (9)a, or the compound may be fully idiomatic, as in (9)b.<sup>18</sup>

- (9) a. lees-moeder  
read-mother  
'mother who is a volunteer reading helper / a mother who reads a lot / a mother to whom you can read a lot / ...'
- b. baar-moeder  
give.birth<sup>19</sup>-mother  
'womb'

As the non-head is adjacent to the head on the surface, there is no direct evidence to postulate an intervening head. In fact, below I will present evidence against such a head and I will conclude the non-head is a bare root.<sup>20</sup> I will therefore refer to this type of primary compounds as root primary compounds or RPCs.

To be entirely clear, the claim in this article is not that all primary compounds in Dutch have a bare root as their non-head. I adopt the view that nominal primary compounds do contain structure which restricts them to nominality. The claim is rather that there is a subtype of primary compounds in Dutch, which I have called root primary compounds, for which a bare root analysis is the only one which is compatible with the data and which as such supports the root-hypothesis.

## 2.3 Evidence against intervening categorial heads ('little heads')

### 2.3.1 Introduction

In this section I argue against the hypothesis that the non-head of a root primary compound first merges with a null lexical categorial head, such as  $v^\circ$  or  $a^\circ$ , before merging with the head (cf. Harley 2009). I point out that the presence of overt categorial heads leads to ungrammaticality. I assume late insertion of vocabulary items (Halle and Marantz 1993). It therefore stands to reason that if the

<sup>18</sup> Of course, all idiomatic compounds may be interpreted literally as well.

<sup>19</sup> It is not entirely clear whether *baar* should be glossed as *give birth*, which is its contemporary meaning, or as *bear*, to which it goes back etymologically.

<sup>20</sup> Such an analysis has been pursued for root compounds with an adjectival non-head by De Belder and Van Koppen (2012). They show that for a specific subtype of AN(N) compounding the alleged A lacks all adjectival properties. As such, it contrasts with compounds which contain a genuine AN-phrase as their non-head. For the first type the alleged A remains uninflected and resists degree modification. The compound is assigned compounding stress. They conclude the alleged A is a root. For the second type they show that the A is inflected and can select degree modification. The AN non-head is assigned DP stress. They conclude the non-head of this type is a partial NP.

presence of overt markers is ungrammatical, the presence of null markers is illicit as well. After all, the vocabulary item is only inserted post-syntactically and its phonological properties should not influence the grammaticality of the structure. This section further contains a discussion of apparent counterexamples. I present examples which seem to contain an overt verbal suffix. I will argue, however, that what could be considered to be a categorial head is part of the root in these examples. I conclude that the non-head of a root primary compound cannot merge with categorial heads.

This section only contains a discussion of  $v^\circ$  and  $a^\circ$ , excluding  $n^\circ$ . Recall that RPCs exist alongside nominal primary compounds (see section 2). We therefore expect that nominal categorial heads can licitly merge on top of the non-head of a primary compound, in which case a nominal primary compound is derived.<sup>21</sup> I will therefore postpone the discussion on nominal categorial heads till section 4 and I will restrict the discussion in this section to verbal and adjectival heads.

### 2.3.2 The ungrammaticality of overt categorial heads

Proponents of Distributed Morphology argue that categorial heads may be either overt, in which case they are overt derivational affixes, or null, in which case they should be understood as null derivational affixes (see for example Embick and Marantz 2008). In this section I argue against the assumption that the non-head of a root primary compound merges with a null categorial head.<sup>22</sup>

Let us assume that RPCs actually contained a null categorial head, as in (10).

- |                                    |                               |                               |                               |
|------------------------------------|-------------------------------|-------------------------------|-------------------------------|
| (10) a. snel-∅ <sub>A</sub> -trein | b. fris-∅ <sub>A</sub> -drank | c. slaap-∅ <sub>V</sub> -pill | d. speur-∅ <sub>V</sub> -hond |
| fast-∅ <sub>A</sub> -train         | fresh-∅ <sub>A</sub> -drink   | sleep-∅ <sub>V</sub> -pill    | track-∅ <sub>V</sub> -dog     |
| 'high-speed train'                 | 'soft drink'                  | 'sleeping pill'               | 'tracking dog'                |

If the representations in (10) were correct, we predict that *overt* categorial heads can be present in RPCs as well. After all, null categorial heads and overt categorial heads are only distinguished at PF given late insertion. Structurally, they are identical. However, this prediction is not borne out. The compounds in (11) and (12) are ungrammatical, despite the fact that compounding is highly productive in Dutch. (11) shows RPCs with intervening overt adjectival suffixes,<sup>23</sup> (12) shows examples with verbalizing affixes.<sup>24</sup> I made sure that the morphologically complex adjectives and verbs in the non-heads are common and acceptable words when occurring independently.<sup>25</sup>

---

<sup>21</sup> I thus propose that all primary compounds of which the non-head contains a nominal affix are instances of nominal primary compounds. Recall that nominal primary compounds contain a linking phoneme. Given that this linking phoneme instantiates a piece of nominal inflection, De Belder (to appear, chapter 3) argues that it can be null. It is common for inflection to be realized as a null morpheme. Examples with a nominal affix are thus examples of nominal primary compounds, even if they do not contain an overt linking phoneme.

<sup>22</sup> In Harley's (2009) approach to compounding in Distributed Morphology, the non-head invariably merges with a categorial head.

<sup>23</sup> Dutch affixes cannot be translated straightforwardly into English.

<sup>24</sup> Compounds with particle verbs as their non-head are perfectly well-formed, e.g. *weg-geef-prijs* 'away-give-price' (give away price, i.e. very low price). I assume they are phrasal compounds: their non-head is a phrase. See Harley (2009) for an analysis of phrasal compounding in Distributed Morphology.

<sup>25</sup> An anonymous reviewer, who is clearly a native speaker of Dutch, doubted whether the prefixes *ont-* and *be-* indeed give rise to ungrammaticality if they are contained in the non-head of a RPC. I have therefore tested twelve neologisms in a questionnaire. The questionnaire contained 18 fillers. Each new compound was introduced by means of a context of which the purpose was to make its use plausible from a pragmatic point of view. I asked the informants whether they could imagine a native speaker of Dutch would use the compound. The questionnaire was responded to by 509 native speakers of Dutch. The scale was a five-point scale with 1 being the score for utter ungrammaticality and 5 being the

- (11) a. \* spaar-zaam-attitude  
save-ADJECTIVAL SUFFIX-attitude

- b. \* eet-baar-datum  
eat-ADJECTIVAL SUFFIX-date
- c. \* vet-ig-dieet  
fat-ADJECTIVAL SUFFIX-diet

- (12) a. \* be-plant-seizoen  
VERBAL SUFFIX-plant-season

- b. \* menstru-eer-pijn  
menstru-VERBAL SUFFIX-pain
- c. \* ont-vlam-tijd  
INCHOATIVE VERBAL SUFFIX -flame-time

Note that the illicitness is not due to the fact that newly formed RPCs are impossible words. Compounding is highly productive and the formations in (13), for example, are fully acceptable.

- |  |   |   |
|--|---|---|
| (13) a. spaar-attitude<br>save-attitude<br>'attitude towards saving' | b. eet-datum<br>eat-date<br>'eating date' <sup>26</sup> | c. vet-dieet<br>fat-diet<br>'diet based on fat' |
|--|---|---|

Admittedly, the illicitness of the examples in (11) and (12) is not necessarily due to the fact that the non-head needs to be a root. I am sure alternative accounts can be proposed to exclude these examples. However, it is not clear to me how their ungrammaticality could be derived without excluding the structures in (10) as well. After all, the examples in (10), (11) and (12) are structurally identical.<sup>27</sup> I conclude that the representations in (10) are wrong. RPCs do not contain an intervening null categorial head.

---

score for complete grammaticality. Yet, the scale was not represented by means of numbers to the informants, but by means of paraphrases of judgments ('sounds very strange' (1), 'sounds rather strange' (2), 'I am not sure' (3), 'sounds more or less OK' (4), 'sounds completely normal' (5)). The average score of the twelve test items was 2,6 on a scale of 1 to 5. I conclude the compounds are far from fully acceptable. Admittedly, the compounds improve if the verb contained in the non-head is part of common Dutch vocabulary and if the root is a trochee, thus improving the rhythm of the compound. The test item *be-tegel-ploeg* 'BE-tile-team' (tiling team, i.e. team which is responsible for tiling), for example, scored 3,3/5 as *betegelen* 'to tile' is a common Dutch verb and *tegel* is a trochee. For the same reasons, *ont-wortel-ploeg* 'ONT-root-team' (team responsible for removing roots of trees) even scored 3,8/5. As such, it was the test item with the highest score. Note that even the best test item on average still does not sound more or less OK to native speakers of Dutch.

<sup>26</sup> The compounds are acceptable, even though it may not be very clear what they might refer to.

<sup>27</sup> The difference between the overt and covert categorial heads only comes at play at vocabulary insertion, i.e. post-syntactically. Syntactically, they are all identical.

### 2.3.3 Apparent counterexamples

In this section I discuss examples which at first sight seem to contradict the claim that the non-head of a root primary compound cannot contain a verbal derivational affix. I argue that the non-heads of the alleged counterexamples are not morphologically complex.

Consider word-forms which include the prefix *ver-* in Dutch. There are four uses which are productive. An overview is given below.

- |   |  |
|---|--|
| (14) a. De kat ver-dik-t.<br>the cat VER-thick-3SG<br>'The cat is fattening.'                             | b. Het Belgische leger ver-vlaams-t.<br>the Belgian army VER-Flemish-3SG<br>'The Belgian army becomes more Flemish.' |
| (15) a. Oude honing ver-suiker-t.<br>old honey VER-sugar-3SG<br>'Old honey crystallizes.'                 | b. De relatie ver-water-t.<br>the relation VER-water-3SG<br>'The relationship is fizzling out.'                      |
| (16) a. Mariever-slaap-t haar tijd.<br>Mary VER-sleep-3SG her time<br>'Mary is wasting time by sleeping.' | b. Mariever-gok-t haar geld.<br>Mary VER-gamble-3SG her money<br>'Mary is wasting money by gambling.'                |
| (17) a. Mariever-slaap-t zich.<br>Mary VER-sleep-3SG self<br>'Mary overslept.'                            | b. Mariever-draai-t de waarheid.<br>Mary VER-twist-3SG the truth<br>'Marie is twisting the truth.'                   |

In (14) the verb with *ver-* refers to an increasing degree. The predicate the degree refers to is expressed by the root. In (15) the verb with *ver-* can be paraphrased as 'grow into the predicate expressed by the root'. The meaning may be metaphoric. In (16) *ver-* indicates that the direct object is wasted by doing the event expressed by the root. Finally, in (17) *ver-* expresses that the result of the event the root refers to is undesirable.

Given that these verbs are derived by means of a productive word-formation process, they are undoubtedly morphologically complex. As the non-head in root primary compounds they behave as expected. Their morphological complexity results in ungrammaticality. This can be deduced from the examples below.

- |   |   |
|---|---|
| (18) a. * ver-dik-probleem<br>VER-thick-problem | b. * ver-vlaams-tendens<br>VER-Flemish-tendency |
| (19) a. * ver-suiker-honing<br>VER-sugar-honey  | b. * ver-water-relatie<br>VER-water-relation    |
| (20) a. * ver-slaap-tijd<br>VER-sleep-time      | b. * ver-gok-geld<br>VER-gamble-money           |
| (21) a. * ver-slaap-pech<br>VER-sleep-bad.luck  | b. * ver-draai-manier<br>VER-twist-manner       |

There is yet a fifth group of verbs which include *ver-*, which is by far the largest group. In these verbs the prefix has no clear, systematic meaning and the word-formation process is improductive. Examples of this type are given in (22).

- |  |  |
|--|--|
| (22) a. Mariever-jaar-t.<br>Mary VER-year-3SG<br>'Mary celebrates her birthday'                  | b. Mariever-taal-t de tekst.<br>Mary VER-language-3SG the text.<br>'Marie translates the text.'      |
| c. Deleerkracht verwen-t Marie.<br>the teacher spoil-3SG Mary<br>'The teacher is spoiling Mary.' | d. Deleerkracht vermaan-t Marie.<br>the teacher reprimand-3SG Mary<br>'The teacher reprimands Mary.' |
| e. Marieverdwijn-t.<br>Marie disappear-3SG<br>'Marie disappears.'                                |  |

Although it is clear that these words are morphologically complex from an etymological point of view, it is hard to determine whether they are indeed complex from a synchronic, morphological point of view. For some verbs this seems plausible, as they are quite transparent, for others this seems less likely. The internal structure of the examples (22)a and (22)b is probably accessible to the native speaker. In contrast, the internal structure of (22)d and (22)e has to be opaque as *maan* is archaic and peripheral and *dwijn* does not have an independent meaning in contemporary Dutch. The verb in (22)c resembles (and is etymologically related to) the verb *wennen* 'to get used to', but a lack of a semantic relation between *verwennen* 'to spoil' and *wennen* 'to get used to' may be confusing the native speaker. However, it turns out that verbs with *ver-* belonging to the improductive group can occur as the non-head of a root primary compound.

- |   |  |   |
|---|--|---|
| (23) a. verjaar-dag<br>celebrate.one's.birthday-day<br>'birthday' | b. vertaal-bureau<br>translate-agency<br>'translation agency'    |   |
| c. verwen-dessert<br>spoil-dessert<br>'rich dessert'              | d. vermaan-brief<br>reprimand-letter<br>'letter with reprimands' | e. verdwijn-truc<br>disappear-act<br>'disappearing act' |

Given the licitness of the examples above, we have to conclude that what has been analyzed as the improductive prefix *ver-* is actually not a prefix, but part of the root. To conclude, if a form which seems to be morphologically complex can occur as the non-head of a root primary compound, this deviant behavior correlates with improductivity.

If one accepts the conclusion, we now have found an independent testing ground to establish whether a certain form is morphologically complex or not. If a form can occur as the non-head of a root primary compound, it is not morphologically complex. This is a significant result. I will present yet another set of examples below.

The examples in (24) repeat the general observation that the affix *-eer* cannot be contained in a root primary compound.

- (24) a. \* menstru-eer-pijn  
menstru-VERBAL SUFFIX-pain

b. \* calcul-eer-programma  
calcul-VERBAL SUFFIX-programm

c. \* pollu-eer-niveau  
poll-VERBAL SUFFIX-level

Yet, the compounds in (25) are fully acceptable.

- (25) a. parkeer-garage  
park-garage  
'parking garage'  
b. kampeer-terrein  
camp-ground  
'campground'

*Parkeer* ‘park’ and *kampeer* ‘camp’ are not the only items ending in *-eer* which can occur as the non-head of a RPC. (26) shows more examples.

(26) accentueer-stift 'marker pen', adresseer-machine 'addressing machine', arceer-lijn 'hatching line', balanceer-kunst 'balancing act', blokkeer-stift 'locking pin', calqueer-papier 'tracing paper', codeer-machine 'coding machine', defileer-mars 'parade march', doseer-dop 'dispensing cap', frankeer-zegel 'postage stamp', kalmeer-middel 'calmative' (Lit. calm-remedy), kampeer-terrein 'campground', markeer-speld 'marker pin', parkeer-bon 'parking ticket', pikeer-mesje 'knife for small plants' (Lit. pick.out-knife), regeer-periode 'period of office' (Lit. govern-period), scalpeer-mes 'scalping knife', soldeer-bout 'soldering iron', sorteer-centrum 'sorting center'

In order to capture the contrast between the data in (24) and the examples in (25)-(26) I propose that the non-heads in (25)-(26) are roots which simply happen to end in *-eer*, whereas the non-heads in (24) genuinely contain the verbal suffix *-eer*. I will first present some properties of the verbal suffix *-eer* and I will then show that the examples in (25)-(26) do not have these properties.

First note that there surely is a verbal affix *-eer* in Dutch. Some more examples are given in (27).



There is little doubt it is indeed a derivational suffix rather than being part of verbal inflection, as it can precede the nominal derivational suffix *-ing*, as is shown in (28). In other words, it occurs within the derivational domain.

- (28) a. abstrah-eer-ing  
abstract-EER-ING  
'abstraction'  
b. conserv-eer-ing  
preserv-EER-ING  
'preservation'  
c. oriënt-eer-ing  
orient-EER-ING  
'orientation'

The suffix *-eer* has some specific properties which are relevant to the discussion. Firstly, it only attaches to Latinate, bound roots. As such, it is in complementary distribution with the nominal suffix *-ie* (surfacing in various forms such as *-atie*, *-entie*, ..., see de Haas and Trommelen 1993) which also selects for Latinate, bound roots. This is illustrated in (29).

(29)	Verb	Noun
	abstrah-eer-en	'to abstract'
	administr-eer-en	'to administrate'
	arrest-eer-en	'to arrest'
	articul-eer-en	'to articulate'
	calcul-eer-en	'to calculate'
	castr-eer-en	'to castrate'
	combin-eer-en	'to combine'
	communic-eer-en	'to communicate'
	abstrac-tie	'abstraction'
	administr-atie	'administration'
	arrest-atie	'arrest'
	articul-atie	'articulation'
	calcul-atie	'calculation'
	castr-atie	'castration'
	combin-atie	'combination'
	communic-atie	'communication'

Secondly, the suffix *-eer* triggers some systematic consonantal changes (De Haas and Trommelen 1993). Most notably, if the final consonant of the root contained within the noun derived by means of *-ie* surfaces as /k/, it will surface as /s/ in the verb derived by *-eer*. As a consequence, a root preceding the suffix *-eer* will never surface with a /k/ as its final consonant.

- (30) a. publiceर /pybliser/  
       publicatie /pyblikasi/<sup>28</sup>
- b. provoceर /provoser/  
       provocatie /provokasi/
- c. communiceer /kɔmyniser/  
       communicatie /kɔmynikasi/

In sum, the verbal suffix *-eer* selects for bound, Latinate roots and as such it is in complementary distribution with the nominal suffix *-ie*. It may further trigger a consonantal change. Because of this consonantal change a root preceding *-eer* will never surface with a /k/ as its final consonant.

In the examples in (25)-(26) the properties of the suffix *-eer* cannot be recognized. Firstly, for none of the items in (26) exists a corresponding noun containing the nominal suffix *-ie*, as can be seen in (31).<sup>29</sup>

---

<sup>28</sup> The phonological transcription represents the Belgian Dutch pronunciation. Nothing hinges on this fact.

<sup>29</sup> Admittedly, I did find about 10 counterexamples in Dutch dictionaries. Interestingly, they belong to jargonese (and the non-head thus refer to a specific submeaning of the verb). Examples are *reageer-buis* 'test tube' (Lit. react-tube) (chemistry) and *separeerkamer* 'padded cell' (Lit. separate-room) (psychiatry) (versus *reactie-tijd* 'reaction time' and *separatie-angst* 'separation anxiety'). I have further found an example in which the non-head expresses a euphemism (*animeer-meisje* 'prostitute' (Lit. animate-girl)). These examples support the hypothesis that jargons may have a distinct (non-natural) grammar (see Jaspers 2009 and references therein) or that otherwise complex structures may be relisted as roots if the root refers to a specific submeaning.

- (31) \*accentuatie, \*adressatie, \*arcatie, \*balancatie, \*blokkatie, \*calquatie, \*codatie, \*defilatie, \*dosatie, \*frankatie, \*kalmatie, \*kampatie, \*markatie, \*parkatie, \*pikatie, \*regatie, \*scalpatie, \*soldatie, \*sortatie

Secondly, the consonant /k/ may precede *-eer* in the examples in (32).

- (32) calqueer /kalker/, blokkeer /bloker/, frankeer /franker/, markeer /marker/, parkeer /parker/, pikeer /piker/

In sum, the items in (25)-(26) do not reflect the properties of the suffix *-eer*. I therefore conclude that the examples in (25)-(26) do not contain the suffix *-eer* at all. The non-heads are simply roots. As such, it should not come as a surprise that exactly the same items may occur as the non-head of RPCs. The fact that they are morphologically simplex licenses their occurrence as non-heads in RPCs.<sup>30</sup>

I conclude that these observations indicate that whenever a form occurs as the non-head of a root primary compound, this form is morphologically simplex. Interestingly, we can conclude that root primary compounds are an ideal testing ground to determine whether a form is morphologically complex or not. As the non-head of such compounds can only contain roots, only underived word-forms can be attested in this syntactic context. We have seen that a form which at first sight might seem morphologically complex can still function as a root. Interestingly, being morphologically simplex correlates with other morphological properties, such as improductivity. It did not correlate with transparency. Less words may be morphologically complex than has been hitherto assumed.

### 2.3.4 Conclusion

In this section I have argued that the non-head of a root primary compound cannot merge with a categorial head. I have demonstrated that the non-head cannot contain an overt derivational affix. I have finally discussed some apparent counterexamples for which I argued they are morphologically simplex. This result is interesting. It turns out that the non-head of a RPC is an ideal testing ground to determine whether a form is morphologically complex or not.

In this section we have seen that verbal or adjectival heads are excluded from the non-head of a RPC. In the next section I address the more complex issue of nominal heads.

---

<sup>30</sup> One may wonder why the roots in the non-heads in (26) then so obviously contain the ending *-eer*. Note in this respect that most roots in (26) (except for *sorerteer*, *frankeer*, and *scalpeer*) are borrowed from French infinitives (see *Woordenboek der Nederlandsche Taal* under each specific lemma). Hence, what one recognizes systematically is a French (or archaic French) infinitive. For each case the French infinitival marker *-eer* became part of the root, resulting in formal systematicity. One might even recognize the French root which is contained in the infinitive. For example, speakers of Dutch will recognize *adres* (address) and *kalm* (calm) which have been loaned independently.

## 2.4 The non-head of a RPC is not nominal

### 2.4.1 Introduction

In section 2 I have argued there are two types of primary compounds in Dutch, viz. nominal primary compounds and root primary compounds. Admittedly, postulating two different compounding types is not the most parsimonious starting point one can imagine. It is therefore tempting to hypothesize the two types can be reduced to a single type. In other words, one could argue that the distinction between nominal primary compounds and root primary compounds is only apparent and that there is only one type of primary compounding in Dutch. For the nominal primary compounds there is little reason to doubt we are on the right track by arguing the non-head is nominalized. It is striking that the set of licit non-heads of this type is fully congruent with the set of attested nouns in Dutch.<sup>31</sup> Yet, one might hypothesize the non-head of a RPC is in fact nominal, thereby postulating that all primary compounds are nominal primary compounds.

There is no overt nominalizing head in RPCs. Yet, it is not difficult to imagine a phonologically null nominalization strategy. One could postulate a zero nominalizing head, i.e. a null  $n^\circ$  (cf. Harley 2009) or a Compounding Frame which nominalizes the non-head by defining the non-head as a nominal position (Borer 2013a). One could even argue a nominalization strategy is superfluous as all roots are nominal *in se* (Kayne 2009, Acquaviva) or because the nominal category is the category which is assigned to a non-categorized root by default. As such, the absence of an overt nominalizer for RPCs does not count as a counter-argument.

Yet, in this section I argue that the hypothesis according to which the non-head of a RPC is nominal cannot be maintained. It is thus not possible to unify both types of primary compounds. I present three arguments. I first point out that the non-head of an RCP does not need to be co-extensive with an existing noun. I then illustrate that the non-head is not necessarily interpreted nominally. I finally show that even the lowest, otherwise obligatory nominal functional projections are absent in RPCs. In each section I contrast RPCs with nominal primary compounds. The non-head of each single nominal primary compound has the same distribution as roots which are licensed in a *bona fide* nominal positions, it is invariably interpreted nominally and it may retain nominal functional projections.

---

<sup>31</sup> The term *noun* refers to a root in a nominal position, not to a lexical category.

### 2.4.2 A distributional contrast between RPCs and nominal primary compounds

There are many roots which surface as the non-head of a RPC, while they are not licit in *bona fide* nominal positions.<sup>32</sup> For example, the non-heads of the RPCs in (33) cannot occur as the lowest projection in a DP, as is shown in (33).<sup>33, 34</sup>

- (33) a. speur-hond      b. achter-deur      c. niet-verjaardag  
           track-dog                  back-door                  not-birthday  
           'tracking dog'                'backdoor'                'unbirthday'

- (34) a. \* de speur      b. \* de achter      c. \* de niet  
           the track                  the back                  the not

Under the assumption that the roots in the non-heads in (33) are nominalized, it is unclear why they do not have the distribution of nouns elsewhere. Under the proposal that they are bare roots, in contrast, it follows immediately. I therefore conclude these data support the view that the non-heads in (33) are bare roots.

Nominal primary compounds contrast sharply with RPCs in this respect. The set of roots which can occur as the non-head of nominal primary compounds is a subset of the roots which can occur as the lowest projection in a DP.<sup>35</sup> This claim is not new. It is common knowledge in the literature on Dutch morphology that the presence of the overt linking elements *-s* and *-en* (the hallmark of nominal primary compounds) implies the nominality of the non-head.<sup>36</sup> The claim can be illustrated

<sup>32</sup> Root-based models overgeneralize as they predict that any root can be used in any syntactic context, a prediction which is clearly false. To the best of my knowledge, an overall solution to this problem has not yet been found, though see Borer (2005b: chapter 11 and 2013:chapter 13) for suggestions.

<sup>33</sup> Admittedly, any vocabulary item can be nominalized by using it to refer to itself, as *not* in *The 'not' on the blackboard*. De Belder and van Craenenbroeck (2013) present five tests to distinguish between regular vocabulary items in a root position and specific cases of self-reference. Most tests cannot be applied to items contained in a compound, yet, De Belder and van Craenenbroeck point out that self-referring items always have a quotative nature. The actual phonetic form should be at stake meaningwise. For example, if *niet* in (33)c had been used self-referentially, it would only refer to the phonetic string /nit/ and it would thus not imply negation. In other words, the meaning of self-referring items differs sharply from the meaning of the non-head in the examples under discussion. Of course, self-referring items can be used in a RPC, as in *frrpl-geluidje* 'frrpl-sound'.

<sup>34</sup> I have noted in section 2 that RPCs are at least as old as the oldest written sources of Dutch. Unsurprisingly, then, the non-head of a RPC may be archaic. The root in the non-head of some RPCs was co-extensive with a noun in an earlier stage of Dutch, but it is no longer co-extensive with a noun in present-day Dutch. Yet, it has been preserved within the RPC. For example, the roots *wal* and *sperzie* ceased to occur as independent nouns. Yet, they still occur as the non-head of RPCs, as in *wal-vis* 'whale' (Lit. whale-fish) and *wal-dieren* 'members of the family of the whale' (Lit. whale-animals) and as in *sperzie-boon* 'green bean' (Lit. asparagus-bean) and *sperzie-torrelje* (specific kind of beetle, lit. *asparagus-beetle*). It is unclear to me whether these words are seen as morphologically complex by the native speaker (see section 3.3 on archaisms and morphological complexity). If they are, they show that the non-head of a RPC does not need to be co-extensive with a noun.

<sup>35</sup> I imagine some roots may be found within a DP, while they cannot be found as the non-head of a nominal primary compound. I therefore suspect the set of roots which can occur in a DP is a superset of those which occur as the non-head of a nominal primary compound.

<sup>36</sup> There is a schwa which may occur in other types of compounding to improve the rhythm of the compound, e.g. *drinkebroer* 'alcoholic' (Lit. drink-E-brother) and *oudejaar* 'New year's Eve' (Lit. old-E-year)(see chapter 5).

by means of derivational suffixes. Derivational suffixes may select an overt linking phoneme, as is shown in (35).<sup>37</sup>

- (35) a. prins-es-en-jurk  
prince-ess-LP-dress  
'gown as worn by princesses'  
b. over-heid-s-instelling  
above-HEID-LP-institution  
'governmental institution'

If a derivational suffix which is contained in a non-head selects an overt linking element it can be predicted the suffix is nominal. In other words, there are no non-nominal suffixes which select an overt linking phoneme if they are part of a compound's non-head. Again, this shows that the non-head of nominal primary compounds is obligatorily nominal. In this respect, they contrast with RPCs.

#### 2.4.3 An interpretational contrast between RPCs and nominal primary compounds

The non-head of a RPC does not need to be interpreted nominally. For example, the compound in (36) is ambiguous between a nominal and a verbal reading for the non-head. This corresponds with the fact that the root in the non-head occurs commonly in both nominal and verbal contexts, as can be seen in (37).

- (36) antwoord-centrum  
answer-center  
'center where answers are stored / center where answers are provided/...' (nominal readings)  
'center where one answers the phone/...' (verbal reading)

- (37) a. ik antwoord      b. het antwoord  
I answer                the answer  
'I answer'              'the answer'

This fact suggests that the nominal nature of the non-head is not syntactically implied.

In contrast, the presence of an overt linking phoneme forces a nominal interpretation. For example, *a priori* one may expect the compound in (38) to be ambiguous between a nominal and verbal reading, as in the RPC above. Yet, the nominal primary compound triggers an exclusive nominal interpretation of the non-head.

---

<sup>37</sup> Some nominal derivational suffixes do not select an overt linking phoneme, such as -(at)ie in *menstruatie-pijn* 'menstrual pain'. It may be the case that these suffixes may be followed by a null linking phoneme (see section 2). Alternatively, one might hypothesize that whatever role is fulfilled by linking phonemes can be fulfilled by these derivational suffixes as well, in which case they are in complementary distribution with linking phonemes. Note in this respect that De Belder (to appear, chapter 3) argued that linking phonemes are class markers. Note further that the suffixes which do not select an overt linking phoneme are Latinate suffixes. Lowenstamm (2007) and Kramer (2009) have argued that loaned morphemes may function as class markers in Jiddisch and Amharic. As such, it is plausible that Latinate suffixes function as class markers in Dutch. Under these assumptions the fact that they are in complementary distribution with the linking phonemes -en and -s then follows immediately. (Note that it is not the case that a loaned suffix cannot surface with an overt linking phoneme. The suffix -eut, for example, which is borrowed from Greek, selects -en, as in *techn-eut-en-conferentie* 'boffin conference'.)

- (38) antwoord-en-centrum

answer-LP-center

'center where answers are stored / center where answers are provided/...' (nominal readings)

\* 'center where one answers the phone/...' (verbal reading)

The fact that the non-head is necessarily interpreted nominally indicates that the nominality of the non-head is syntactically implied for nominal primary compounds. I conclude the data show that the non-head of the RPC is not nominal structurally, whereas it is syntactically nominal in nominal primary compounds.

#### 2.4.4 A structural contrast between RPCs and nominal primary compounds

##### 2.4.4.1 Introduction

In this section I show that RPCs and nominal primary compounds differ in the fact that the non-head of a RPC is invariably bare, whereas the non-head of a nominal primary compound may contain nominal inflection. In the first subsection I show that even the lowest, most expected nominal projections are absent on the non-head of a root primary compound. In the second subsection I contrast RPCs with nominal primary compounds which may contain nominal inflectional elements. In a third subsection I provide further support for the view that the empirical contrasts observed in the previous two subsections indeed should be ascribed to the fact that there are two different compounding types.

##### 2.4.4.2 The non-head of a RPC cannot contain nominal inflection

In this section I show that even the most expected nominal inflectional markers are absent on the non-head of a RPC. More specifically, I discuss three instances of inflection from Dutch, viz. gender marking in West-Flemish, number marking on pluralia tantum and a remnant of an archaic class marker on dialectal neuter nouns. I will finally discuss a Swedish inflectional marker. For these projections it has been argued that they are very low in the nominal structure. In fact, it has been claimed they can even be identified as the nominal categorial head, i.e.  $n^o$ , which attaches directly to the root (see Lowenstamm 2007 for gender as  $n^o$ , see Acquaviva 2008 for plural marking on pluralia tantum as  $n^o$ , see Kihm 2005 for class markers as  $n^o$ ).

First consider gender marking. In Standard Dutch the gender of a noun can only be read off from agreement in the DP (see Schoorlemmer 2009). In contrast, West-Flemish dialects have overt gender marking on the noun. More specifically, dialects which belong to this group have overt gender marking on feminine nouns by means of a schwa ending (Haegeman 1998). This is illustrated in (39).<sup>38</sup> (39)d shows that this gender marking is obligatory in a DP.<sup>39</sup>

- |   |   |   |                         |
|---|---|---|-------------------------|
| (39) a. e vrouw-e<br>a woman-F<br>'a woman' | b. e stroat-e<br>a street-F<br>'a street' | c. e school-e<br>a school-F<br>'a school' | d. * e vrouw<br>a woman |
|---|---|---|-------------------------|

---

<sup>38</sup> The West-Flemish examples in this article are from Blankenberge Dutch. I would like to thank Katlijn Van Audenaerde and Monica Roose for data and judgments.

<sup>39</sup> Masculine and neuter nouns lack this ending, e.g. *vent* 'man' (masculine) and *kind* 'child' (neuter).

In RPCs this gender marking is absent, as shown in (40).<sup>40</sup>

- |   |   |  |
|---|---|--|
| (40) a. vrouw-mens<br>woman-human<br>'woman (pejorative)' | b. stroat-kat-e<br>street-cat-F<br>'street cat' | c. school-gebouw<br>school-building<br>'school building' |
|---|---|--|

The non-heads of these compounds do not contain gender marking. We have seen above that it is an obligatory ending on a feminine noun in West-Flemish. I conclude that the compound's non-head in (40) is smaller than the noun. It is a bare root.

Secondly, let us take a look at pluralia tantum. Some examples are given below.

- |   |   |
|---|---|
| (41) a. hersen-en<br>brain-PL<br>'brains' | b. kleer-en<br>cloth <sup>41</sup> -PL<br>'clothes' |
|---|---|

Number marking on pluralia tantum is lexically selected by the root. It is obligatorily present on the noun to derive the intended meaning. In its absence, the noun is either non-existent, as in (42)a, or has a different meaning, as in (42)b.

- |   |  |
|---|--|
| (42) a. * de/het hersen<br>the <sub>COMMON</sub> /the <sub>NEUTER</sub> brain | b. het kleed <sup>42</sup><br>the <sub>NEUTER</sub> dress<br>'the dress' |
|---|--|

Interestingly, the plural markers are absent in RPCs. Yet, the meaning which is contained in the compound refers back to the meaning of the pluralis tantum.

- |   |   |
|---|---|
| (43) a. hersen-helft<br>brain-half<br>'cerebral hemisphere' | b. kleer-kast<br>cloth-closet<br>'wardrobe' |
|---|---|

What is contained in the non-head of these compounds is therefore again smaller than the noun itself. I conclude it is a bare root.

Some neuter nouns in dialects from East-Flanders, West-Flanders and Zeeland may contain a schwa-marking. The examples which are given below are from Blankenberge Dutch.<sup>43</sup>

---

<sup>40</sup> Nominal primary compounds retain gender inflection (see chapter 3). For example, the feminine noun *sirope* 'syrup' keeps its ending in *siropeflasche* 'syrup bottle'. See section 4.4.4.

<sup>41</sup> The English root *cloth* may head a mass DP. The gloss therefore wrongly suggests that *kleer* functions as *cloth*. It does not. *Kleer* is only attested as the root of the pluralium tantum *kleren* 'clothes' or as the non-head of a root primary compound.

<sup>42</sup> The plural *kleren* is a contracted version of the archaic form *klederen*, which is the plural form of *kleed*.

<sup>43</sup> The dictionary *Woordenboek der Nederlandsche Taal* mentions the gender of *oog* 'eye' and *oor* 'ear' may vary between neuter and feminine, De Bo (1892) mentions they are feminine in West-Flemish. Yet, native speakers of West-Flemish whom I have consulted (Katlijn Van Audenaerde, Jan Van Coillie, Lut Lams and Monica Roose) all agree they select neuter articles. Liliane Haegeman (p.c.) assigns feminine gender to *oor* 'ear' and accepts both genders for *oog* 'eye'.

- (44) a. het hert-e      b. het oor-e  
       the<sub>NEUTER</sub> heart-ə      the<sub>NEUTER</sub> ear-ə  
       'the heart'                  'the ear'
- c. het oog-e      d. het strang-e  
       the<sub>NEUTER</sub> eye-ə      the<sub>NEUTER</sub> beach-ə  
       'the eye'                  'the beach'

The schwa-ending on these nouns is a remnant of an ancient inflectional class marker (Freek Vandevelde p.c.). It is not clear whether it has any role in present-day Dutch, yet it is obligatorily present in those dialects which retained it, as can be seen in (45).

- (45) a. \* het hert  
       the heart

Interestingly, it is absent in RPCs, as can be seen in (46). The examples are again from Blankenberge Dutch.<sup>44</sup>

- (46) a. hert-zeer      b. oor-dokteur      c. oog-lenze      d. strand-cabine<sup>45</sup>  
       heart-ache      ear-doctor      eye-lens      beach-cabine  
       'heart ache'      'ear doctor'      'contact lens'      'cabana'

Finally, consider Swedish compounds. As in Dutch, there are Swedish primary compounds which select an overt linking element. The linking element can only be used if the non-head is nominal (Holmberg 1992:32, Josefsson 1997:62,64). It is mostly -s, as in (47).

- (47) a. moder-s-mjölk      b. kök-s-dörr  
       mother-LP-milk      kitchen-LP-door  
       'breast milk'      'kitchen door'

Swedish has strong and weak nouns. Weak nouns are nouns which end in -e or -a in the nominative singular, as in (48).

- (48) a. kvinn-a      b. gat-a  
       woman-INFL      street-INFL  
       'women'      'street'

This ending may be dropped in primary compounds. This is illustrated in (48). (The examples are taken from Josefsson 1997:64.)

<sup>44</sup> The marker is not obligatorily absent. If it is present on the non-head, as in *hert-e-klop* 'heart beat', we are dealing with nominal primary compounds (see below).

<sup>45</sup> Unsurprisingly, there is lexical variation in this domain. Liliane Haegeman never omits the schwa of *strange* in compounds, yet my informants from Blankenberge were very clear on the fact that the schwa should be dropped in (46)d.

- |                    |                |
|--------------------|----------------|
| (49) a. kvinn-folk | b. gat-lykta   |
| woman-people       | street-light   |
| 'women'            | 'street light' |

The non-head of these primary compounds is thus smaller than the noun. Josefsson (1997, p.c.) argues the non-head is a root in these cases. I assume they are RPCs.

In sum, the non-head of a RPC lacks nominal inflectional markers. I conclude this follows immediately from the fact that the non-head is a bare root structurally.

#### *2.4.4.3 The non-head of a nominal primary compound does contain inflection*

In the previous section we have seen that the non-head of a RPC does not contain any inflectional elements. However, inflectional markers are not banned in general from the non-heads of primary compounds. The same roots which occurred bare in the previous section may select an inflectional element as well as the non-head of a primary compound. I argue that in these cases they are part of nominal primary compounds.<sup>46</sup>

First consider feminine roots in West-Flemish. They do not only occur bare as the non-head in RPCs, but they may occur with inflection as the non-head in nominal primary compounds as well, as in (49) (compare the examples with the examples in (40)).

- |                          |                    |                         |
|--------------------------|--------------------|-------------------------|
| (50) a. vrouw-e-klooster | b. stroat-e-loper  | c. school-e-gemeenschap |
| woman-INFL-convent       | street-INFL-runner | school-INFL-community   |
| 'convent for nuns'       | 'tramp'            | 'network of schools'    |

Secondly, consider the roots which are part of *pluralia tantum*. In nominal primary compounds they may select an inflectional marker, as can be seen in ((50)) (cf. (43)).<sup>47</sup>

- |                         |
|-------------------------|
| (51) a. kleer-en-winkel |
| cloth-INFL-shop'        |
| 'clothes shop'          |

Thirdly, neuter nouns which select a schwa marking in dialects from East-Flanders, West-Flanders and Zeeland may select an inflectional marking as the non-head in primary compounds in these dialects as well, as in the examples in (51) from Blankenberge Dutch.

---

<sup>46</sup> The reader may have observed that linking phonemes, gender markers, plural markers and inflectional markers in Dutch are all /ə/ or /ən/. Due to n-deletion and considerable variation in the pronunciation of the string /ən/, the difference between /ə/ and /ən/ is further often hardly audible. I am therefore facing a methodological problem. I can observe that the non-head in nominal primary compounds selects an inflectional element, i.e. a schwa, but I cannot establish the exact nature of the schwa. I therefore gloss all inflectional markers in this section by means of INFL, which is short for inflectional marker. The observations in section 4.4.4 indicate the markers in (50) are most likely gender markers, though.

<sup>47</sup> I was not able to find an example with *hersenen* 'brains'. I guess it is disfavored as the non-head of a compound because it contains a sequence of schwas.

- (52) a. hert-e-klop  
           heart-INFL-beat  
           ‘heart beat’
- b. strang-e-café  
       beach-INFL-café  
       ‘beach bar’

Fourthly, Swedish weak nouns may contain a linking element as well when they are the non-head in a primary compound. The linking element which is selected by a weak noun is not the regular linking element *-s*, but *-o* or *-u*, as in (52). These linking elements are remnants of the weak noun’s genitival marker (Josefsson 1997). (The examples are taken from Josefsson 1997:61, 64.)

- (53) a. kvinn-o-dräkt  
           woman-LP-dress  
           ‘woman’s dress’
- b. gat-u-korsning  
           street-LP-junction  
           ‘crossing’

To summarize, the non-head of a nominal primary compounds may contain nominal inflectional markers which are absent on the non-heads of RPCs. This shows there is a structural difference between the two compounding types. The non-head of the RPC is a bare root, whereas the non-head of the nominal primary compound contains nominal inflection.

#### 2.4.4.4 On null linking phonemes

In the two previous sections I have argued that compounds with an overt inflectional marker, as the compounds containing a linking phoneme in (53), instantiate nominal primary compounds, whereas the ones without such an overt inflectional marker, as in (54), are RPCs. I have shown that one and the same root may occur in both types of compounds.

- (54) a. kat-en-staart            b. kat-en-luik  
       cat-LP-tail                cat-LP-flap  
       ‘cat tail’                ‘cat flap’

- (55) a. kat-god                b. kat-fret  
       cat-god                    cat-ferret  
       ‘cat god’                ‘ringtail’

One might argue the data in (53) and (54) have been misinterpreted. One might postulate the examples do not involve variation in compounding types, but rather variation in linking elements. This alternative hypothesis would then imply that the examples in (54) contain a null linking element. Indeed, if the linking elements in nominal primary compounds realize nominal inflection, one may hypothesize that they are null. It is after all common for inflection to be realized by means of a null morpheme. In this section I will argue that there is indeed a null linking element alongside the overt ones *-s* and *-en*. Most importantly, however, I argue this null linking element is not present in the root primary compounds in (54).

Interestingly, in Standard Dutch roots which refer to mass concepts never select an overt linking element. An example of such a root is *zand* ‘sand’, as in (55).

- (56) a. zand-kasteel                    b. zand-bak  
           sand-castle                      sand-box  
           'sand castle'                  'sandbox'

In Dutch dictionaries one does not find a single instance of a compound of which the non-head *zand* selected either one of the overt linking elements, i.e. *-en* or *-s*. I will refer to these roots as the *zand*-roots. Some more examples of such roots are given in (56).

- (57)     a. kaas-boer                    b. zand-grond                    c. vuur-steen  
           cheese-farmer                sand-soil                        fire-stone  
           'cheese farmer'               'sandy soil'                    'flint'  
  
       d. siroop-fles                    e. klei-grond                    f. wol-fabriek  
           syrup-bottle                clay-soil                        wool-factory  
           'syrup bottle'               'clay soil'                      'wool factory'

One may hypothesize the systematic absence of an overt linking phoneme should be ascribed to the fact that *zand*-roots are illicit in nominal primary compounds and are only contained in RPCs. Yet, such a proposal seems counterintuitive. We have reasons to assume the *zand*-roots in (56) occur in a nominal position. *Zand*-roots are a subset of the roots which may occur in *bona fide* nominal positions. In other words, *zand*-roots systematically may be used as nouns as well, as in (57).

- (58)     a. de        kaas                    b. het        zand                    c. het        vuur  
           the        cheese                the        sand                    the        fire  
           'the cheese'                    'the sand'                    'the fire'  
  
       d. de        siroop                    e. de        klei                    f. de        wol  
           the        syrup                the        clay                    the        wool  
           'the syrup'                    'the clay'                    'the wool'

Alternatively, one may hypothesize the compounds in (56) invariably lack an overt linking phoneme as *zand*-roots systematically select a null linking phoneme. This hypothesis seems to be on the right track. An indication that *zand*-roots select a null morpheme comes from their behavior in West-Flemish.

First note that, as in Standard Dutch, *zand*-roots are not followed by an overt marker in West-Flemish if they are the non-head of a primary compound. However, crucially, this fact only holds for non-feminine mass nouns in this dialect, as can be seen in (59). (58) shows that the roots contained in the non-heads are indeed not feminine, as they do not select a feminine gender marker.

- (59) a. de        koas-(\*)e                    b. het        zand-(\*)e                    c. het        vier-(\*)e  
           the        cheese-(\*)F                the        sand-(\*)F                    the        fire-(\*)F  
           'the cheese'                            'the sand'                            'the fire'

- (60) a. koas-boer                    b. zand-groend                    c. vier-stēen  
       cheese-farmer                  sand-soil                          fire-stone  
       'cheese farmer'                'sandy soil'                      'flint'

In contrast, if the root which refers to a mass concept is feminine, as in (60), its schwa ending is retained within the compound, as in (61).<sup>48</sup>

- (61) a. de siroop-e                b. de kliet-e                c. de sjet-e                d. de col-e  
       the syrup-F                    the clay-F                    the wool-F                    the glue-F  
       'the syrup'                    'the clay'                    'the wool'                    'the glue'

- (62) a. siroop-e-flasche            b. kliet-e-groend            c. sjet-e-fabriek            d. col-e-fabriek  
       syrup-F-bottle                clay-F-soil                    wool-F-factory            glue-F-factory  
       'syrup bottle'                'clay soil'                    'wool factory'                'glue factory'

The fact that gender marking is present in these examples, indicates that the root is followed by nominal inflection in general. Hence, the non-head in these compounds is not a bare root, but a root under nominal inflection. If we do not want to multiply the assumed types of compounding in Dutch beyond necessity, it stands to reason that West-Flemish compounds containing a feminine mass noun followed by a gender marker as their non-head are nominal primary compounds. I extend this conclusion to the non-feminine *zand*-roots in West-Flemish and to the *zand*-roots in Standard Dutch. I propose that compounds with a root referring to a mass concept contain nominal markers, albeit covertly. I therefore postulate the presence of a null linking element for *zand*-roots. In sum, the linking element may be null. Yet, when the linking element is null, nominal primary compounds may still project nominal inflectional markers such as gender marking.

If the compounds in section 4.4.2 contained a null linking element, on a par with *zand*-roots, one predicts they may still select other nominal markers in West-Flemish when they occur as the non-head of a primary compound. However, section 3.4.2 emphasized on the fact that they do not select such markers. We have seen, for example, that the feminine roots *straat* 'street' en *school* 'school' do not select nominal inflection in RPCs, as is shown in (63).

- (63) a. stroat-kat-e                b. school-gebouw  
       street-cat                    school-building  
       'street cat'                'school building'

The absence of any nominal marker in (62) is telling. Even if these examples contained a null linking phoneme, nothing should prevent the non-head from selecting a gender marker if it were truly nominal. I therefore submit the non-heads of the examples in (62) are not nominal, but bare roots. In sum, the examples in (54) are nominal primary compounds, whereas the ones in (62) are RPCs. I thus conclude that the variation which is shown through the contrast between (54) and (62) does not involve variation in linking elements (overt versus null), but rather variation in compounding

---

<sup>48</sup> Note that, as expected, the schwa is not present in Standard Dutch examples, which lack overt gender marking on the noun (e.g. *siroop-fles* 'syrup bottle' and *klei-grond* 'clay soil').

types. The examples in (54) are instances of nominal primary compounds, whereas those in (62) are root primary compounds.

#### 2.4.4.5 Conclusion

The non-heads of RPCs systematically lack inflectional markers, whereas inflectional markers are present on the non-heads of nominal primary compounds. It is further clear that it is not the case that the absence of an overt inflectional marker in RPCs can be ascribed to a null linking phoneme. I have discussed a class of roots which select null linking elements and which still may select gender marking. I conclude RPCs differ from nominal primary compounds structurally. Whereas the former contain a bare root as their non-head, the latter contain a non-head which is nominalized syntactically by means of nominal functional projections.

### 2.4.5 Conclusion

In this section I have discussed three contrasts between RPCs and nominal primary compounds. I first showed that the non-head of a RPC may contain roots which are not attested in *bona fide* nominal positions, such as a DP, whereas nominal primary compounds contain exactly those roots which are attested in DPs. In other words, there is a distributional difference between roots which may be found in RPCs and those which may be found in nominal primary compounds. I have then pointed out that the non-head of a RPC does not necessarily trigger a nominal interpretation, whereas the non-head of a nominal primary compound is only subject to nominal denotations. In sum, the compounding types differ meaningwise. Thirdly, I have shown there is a structural difference between RPCs and nominal primary compounds. RPCs do not contain nominal inflectional markers, whereas nominal primary compounds do contain such markers. I conclude that the absence of nominal properties for the non-head of the RPC follows immediately if one does not assume the non-head of this compounding type is nominal. They follow automatically under the proposal that the non-head of a RPC is a bare root.

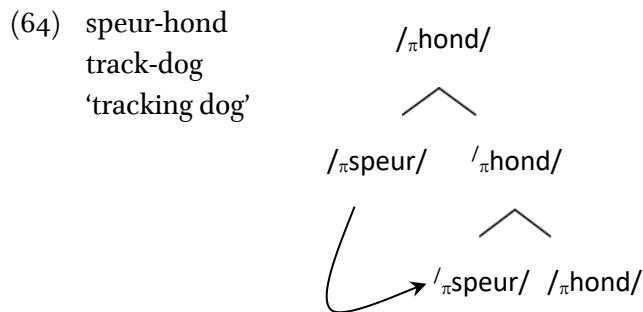
## 2.5 The structure of RPCs

### 2.5.1 Against a syntactic approach

We have seen that the non-head of a root primary compound cannot contain any functional or categorial heads. Empirically, they just seem to contain bare roots and nothing else. Consequently, the structure of a root primary compound should be understood as a direct merger of (phonological indexes of) roots. Borer (2013a:chapter 6) proposes exactly such a structure for primary compounds. Adopting bare phrase structure she proposes that indexes of roots merge directly with one another. The non-head incorporates into the head of the compound for reasons of linearization. The root which projects is the head of the structure. It will be categorized eventually by the functional superstructure. This is illustrated in (63).<sup>49</sup>

---

<sup>49</sup> The  $\pi$  symbolizes that what has been merged is a phonological index of a root.



Borer (2013a:chapter 6, fn 9) submits that the non-head position in the resulting structural configuration is a position which renders the head occupying it equivalent to a noun, at least in English. It is clear that this last aspect of her proposal is undesirable for the root primary compounds in Dutch (see section 4). It thus might be the case that the categorial effects of the structure varies cross-linguistically.

The structure in (63) is attractive for various reasons. It licitly allows a direct merger of roots and it expresses the long-standing insight that root primary compounds involve incorporation (see Mithun 1984, 1986, Baker 1988). However, it faces an immediate problem. RPCs may be recursive, as can be seen in (64).

- |   |   |
|---|---|
| (65) a. fris-drank-automaat<br>fresh-drink-vending.machine<br>'vending machine for soft drinks' | b. was-beer-val<br>wash-bear-trap<br>'raccoon trap'               |
| b. niet-verjaar-dag<br>not-celebrate.one's.birthday-day<br>'unbirthday'                         | c. binnen-speel-tuin<br>inside-play-garden<br>'indoor playground' |

It is not clear how head incorporation can be applied two more than two roots in bare phrase structure. When two minimal projections incorporate, the resulting structure is by definition not-minimal and thus ceases to be a head. Hence, the recursive compound cannot be derived via head movement into a third root. One further wrongly predicts that if a phrase merges with a head, the phrase will be interpreted as the complement of the head (Johnson 2002). To add insult to injury, a RPC may even contain subconstituents, as in (65). By definition it is impossible to incorporate a constituent.

- (66) a. [[bruin-vis]-[baar-moeder]]  
 brown-fish-give.birth-mother  
 'womb of a porpoise'
- b. [[schuif-deur]-[kleer-kast]]  
 slide-door-cloth-closet  
 'wardrobe with a sliding door/sliding doors'

- c. [[[lawine-[zoek-hond]]-[lei-band]]-doos]  
avalanche-search-dog-guide-band-box  
'box to store the leash of an avalanche search dog'

There are ways to circumvent this problem.<sup>50</sup> One could assume that one first builds each subconstituent in a separate derivational workspace. The subconstituent is subsequently spelled out and relisted as a root in Vocabulary (see Harley 2009 on quotative compounds). It will now be able to occupy a root position in the eventual structure. Hence, it can be used as a minimal node and it can incorporate.

Alternatively, one might adopt Carnie's (2000) proposal according to which the terminal stipulation is eliminated from bare phrase structure. As such, phrases and heads are no longer structurally defined. Resultatively, adjunction of a phrase to a head is not excluded and recursive RPCs may be derived by means of incorporation. However, even in this more liberal approach it is still not clear to me how the constituency of (66)a, for example, can be derived successfully via incorporation.

Note further that both proposals overgenerate. One predicts that just any phrase may incorporate. In other words, just any phrase should be able to become the non-head of a primary compound. Clearly, this expectation is not borne out. We have seen in section % that phrases containing verbal heads and adjectival heads are explicitly ruled out. More generally, one still fails to capture the observation that there are precisely two types of primary compounds, viz. RPCs and NPCs. In short, we still fail to understand the essence of what is going on.

### **2.5.2 In favor of a PF approach**

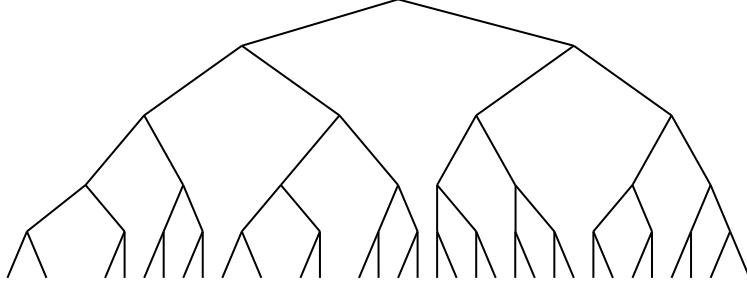
We have seen above that RPCs may be recursive and that they may contain subconstituencies. As such, they cannot be derived via incorporation without violating syntactic principles. From the mere fact that the design of recursive RPCs is at odds with syntactic principles I conclude that RPCs are not built in syntax. I propose they are derived through the post-syntactic fission (Noyer 1992, 1997, Halle 1997) of the root.

Syntax simply builds a structure which contains a single root terminal node. This node is categorized through the functional structure with which it merges (Borer 2005a,b, 2013a). Morphology, a post-syntactic module (Halle and Marantz 1993), splits the node into two separate nodes. This operation is recursive and may split each resulting node into two separate nodes. Each newly derived node is a copy of the previous one, i.e. it is a new root. As such, one post-syntactically derives a fractal-like tree of root terminal nodes from a single syntactic position, as shown in (67).

---

<sup>50</sup> Carnie (2000) proposes to eliminate the terminal stipulation from bare phrase structure. As such, phrases and heads are no longer structurally defined. Nodes can show head-like or phrase-like behavior qua features and movement. Hence, we may conceive them as heads or phrases, but these notions have no theoretical status in syntax. It is thus even possible for a given node to show mixed behavior. Resultatively, adjunction of a phrase to a head is not excluded and recursive RPCs may be derived by means of incorporation. However, even in this more liberal approach it is still not clear to me how the constituency of (65)a, for example, can be derived successfully via incorporation. The system further overgenerates. One predicts that just any phrase may incorporate. In other words, just any phrase should be able to become the non-head of a primary compound. Clearly, this expectation is not borne out.

(67)



The *PF* structure in (67), for example, can host 32 roots, although there is only a single *syntactic* root position. As such, one may insert more than one vocabulary item in the root position post-syntactically.

I discuss the mechanism in detail at a later point. Note first that the present proposal has three immediate advantages. First, one immediately derives that RPCs contain nothing but roots - that is, it follows immediately that they cannot contain any intervening functional projections or that they are not restricted to one category in terms of meaning. Second, it is a welcome result that at narrow syntax the structure contains only a single root position. As such, RPCs are still in compliance with insights on the syntax of roots. Specifically, it follows that *at syntax* each root position is categorized (Arad 2005, Marantz 2008, Borer 2013a), that each derivational workspace contains exactly one root (see De Belder and van Craenenbroeck 2015), and that nominal primary compounds are the only type of compounds which can be derived in syntax (Borer 2013a).<sup>51</sup> Finally, it follows that RPCs are language-specific, as the possibility to fission a morpheme is language-specific and morpheme-specific as well (Noyer 1992, 1997, Halle 1997).

To show how fission may derive a RPC, I present a sample derivation. Suppose Vocabulary Insertion needs to realize the RPC in (68). It has selected six vocabulary items to do so, as can be seen in (69).

- (68) [[[lawine-[zoek-hond]]-[lei-band]]-doos]  
avalanche-search-dog-guide-band-box  
'box to store the leash of an avalanche search dog'

- (69) a. lawine ↔ ✓  
b. zoek ↔ ✓  
c. hond ↔ ✓  
d. lei(d) ↔ ✓  
e. band ↔ ✓  
f. doos ↔ ✓

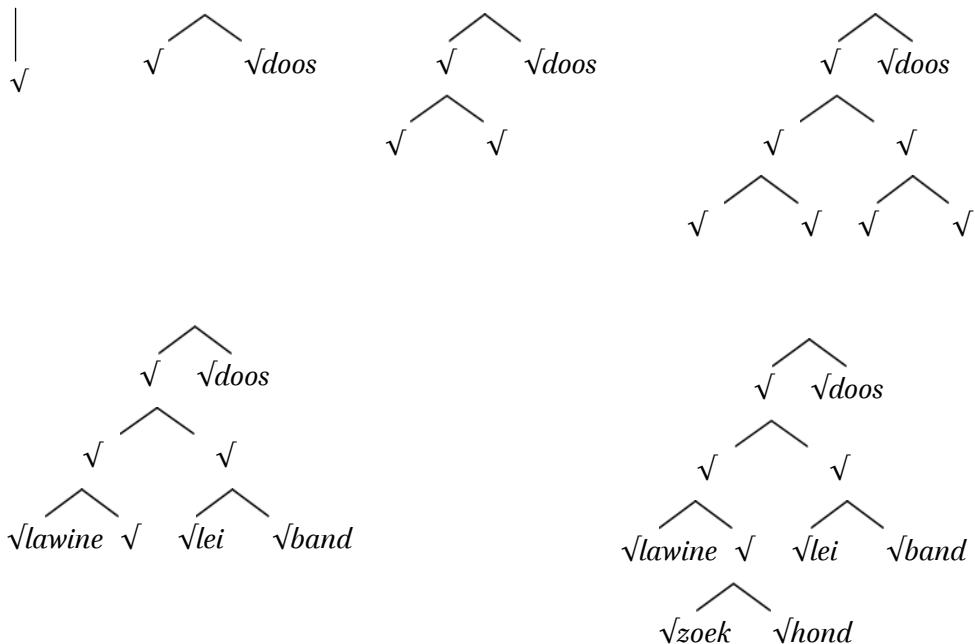
Let us assume that supplementary roots are created at morphology when needed (see Noyer 1997). Consequently, a root morpheme in Dutch may undergo fission at *PF*. Fission is applied when more positions are needed. For example, at *PF* the six vocabulary items in (69) should be inserted, whereas Syntax created only a single root terminal node. Morphology detects that there is only a

---

<sup>51</sup> The idea that two types of compounds can be distinguished on the basis of their syntactic complexity – in particular the idea that in one type we are dealing with a syntactically simplex structure that corresponds to a phonologically complex one- has a precursor for verbal compounds (Ackema 1999).

single root terminal node and it inserts a first vocabulary item, as shown in (70)a. However, five items remain to be inserted. It therefore fissions the node. A subsidiary root is created, as in (70)b. Vocabulary insertion inserts a second vocabulary item. Vocabulary insertion failed to insert the remaining four vocabulary items, and it fissions a node. The newly created node may host a vocabulary item. I assume that for each node fission is optional, but the fission applies until all vocabulary items are inserted.

(70) a.



As such, any subconstituency may be derived. One may wonder how a blind process such as Vocabulary Insertion knows when to fission and what to insert when. I suppose that the operation is ignorant and is a trial-and-error-based mechanism. It creates a structure and subsequently sends it to Encyclopedia. If the structure makes sense at Encyclopedia, it is kept.<sup>52</sup> If it does not, Vocabulary Insertion tries again.

Under the assumption that the order of insertion determines headedness and linear order, they can be read off from the structure. The head is contained in the mother cell. It is interpreted as the head, its gender values are used for phi-agreement in the DP, and it is categorized through its superstructure. This view is in line with the observation that there is always a primary item amongst items realizing a fissioned structure (Halle 1997). I leave a principled approach to these issues for further research.

I assume that the mother cell inherits and reflects the syntactic status of the single root terminal node in syntax. Because the daughter cells are not present in syntax, I assume that they do not have a syntactic status whatsoever. As such, they do not have a categorial status, either.

Previous work on fission exclusively addressed functional terminal nodes. For terminal functional nodes, the supplementary node provides a position to realize features that were not

---

<sup>52</sup> Heidi Harley (p.c.) points out that if Encyclopedia is able to interpret this structure, it follows that its input is not the syntactic structure but the morphologically altered structure under this approach. In other words, the direct connection between PF and LF should be maintained as was assumed in the original model in Distributed Morphology (see Harley and Noyer 1999 and chapter 1).

matched by insertion at the first node according to the Subset Principle. For example, suppose a language contains the vocabulary items in (71). Suppose further (72) is a functional terminal node which syntax created in this language.

- (71) a. /i/  $\leftrightarrow \{a,b\}$   
b. /s/  $\leftrightarrow \{c,d\}$

- (72) Functional morpheme: {a,b,c,d}

If the functional terminal node is fissed at PF, both /i/ and /s/ may be inserted. As such, all features of the functional morpheme will be realized. Given that the saturation of features is involved in fission, one might conclude it belongs to the realm of functional morphemes. However, fully satisfying the realization of functional morphemes does not trigger fission in languages. Functional nodes simply do not trigger fission each time fission might have been the most optimal choice (Halle 1997:137). Whether fission occurs is simply a morpheme-specific and language-specific peculiarity.

Fission is motivated by economy. The learner will prefer to assume fission in order to be able to use a sequence of already existing vocabulary items rather than to postulate a new vocabulary item (Noyer 1997, Halle 1997). Suppose the speaker has learned the morphemes with the feature set {a,b} are realized as /i/ and those with the feature set {c} as /s/. Suppose further that the learner has independent evidence that the functional terminal node {a,b,c} is indeed a single morpheme rather than a sequence of two separate ones (e.g., because of the number of specifiers). The learner will then prefer to analyze the string /is/ for the feature set {a,b,c} as a composition of the exponents /i/ and /s/ rather than storing it as a new vocabulary item.

Applied to RPCs fission prevents the learner from coining a new vocabulary item for each RPC. For example, a learner will prefer to fission a root in order to spell out the PRC *speurhond* ‘tracking dog’. As such, they can recycle the independently existing vocabulary items *speur* ‘track’ and *hond* ‘dog’, and they do not need to postulate the new vocabulary item *speurhond* ‘tracking dog’. In short, the ratio behind fission applies to functional morphemes and to roots equally. I therefore do not see a principled reason to restrict fission to functional morphemes. Thus, I conclude that RPCs are built postsyntactically through fission of the root terminal node.

## 2.6 Theoretical consequences and conclusion

In this article I have distinguished between two types of primary compounds in Dutch. The first type, the nominal primary compound, contains a non-head which is nominalized through nominal functional projections. The second type, the root primary compound, has a bare root as its non-head. I showed that the root in the non-head of the root primary compound cannot merge with a categorial head (i.e. a derivational affix). I further illustrated at length it is not nominalized. The non-head did not show the same distribution of roots as *bona fide* nominal positions, it is not necessarily interpreted nominally and it cannot contain any nominal projections such as gender marking and other inflectional markers. In these respects root primary compounds contrast sharply with nominal primary compounds. An account which captures why Dutch primary compounds are limited to exactly these two types has to await further research.

The non-head of a root primary compound cannot contain any functional or categorial heads. Empirically, they just seem to contain bare roots and nothing else. Consequently, the structure of a root primary compound should be understood as a direct merger of (phonological indexes of) roots (cf. Borer 2013a: chapter 6).<sup>53</sup><sup>54</sup>

The proposal entails a methodological advantage. Given that the non-head of root primary compounds are bare roots we now consist of an empirical testing ground to study the behavior of lexical projections. I have shown that this syntactic context can be used to determine whether a form is morphologically complex or not. If it can occur as the non-head of a root primary compound, i.e. as a root, it is necessarily morphologically simplex.

The proposal has several theoretical consequences. By showing that one can find examples of primary compounds of which the non-head is a bare root this article provides empirical support for the root-hypothesis (Halle and Marantz 1993, Borer 2005a,b, 2013a). It further entails that roots occupy a slot in the structure and that they do not depend on the presence of functional or categorial heads in order to be interpreted (Aarad 2005, Marantz 2007) or to be realized (Ramchand 2008, Starke 2009). Neither are they a modifier of functional structure (Harley 2005, Den Dikken 2008). In short, in this article I have shown that there is a syntactic domain which consists of nothing but bare roots. It supports the existence of roots and it implies roots occupy a node in the structure.

---

<sup>53</sup> Borer (2013:chapter 6) proposes a slightly more elaborate structure for primary compounds. She assumes bare phrase structure and she proposes that indexes of roots merge directly with one another. The non-head incorporates into the head of the compound for reasons of linearization. The head projects. Yet, it is not entirely clear how head incorporation can be applied two more than two roots in bare phrase structure (see also Borer chapter 6, section 4 and chapter 10 for discussion). Borer (chapter 6, fn 9) further submits that the non-head position in the resulting structural configuration is a position which renders the head occupying it equivalent to a noun, at least in English. It is clear that this last aspect of her proposal is clearly undesirable for the root primary compounds in Dutch (see section 4). It is not clear to me why the categorial effects of the structure varies across languages.

<sup>54</sup> Note that such a structure may be recursive. This result is desirable. It is not the case that root primary compounds have to be bipartite (see for example (5)h in which one may recognize three roots (*niet, verjaar, dag*)).

## 3 Linking Elements

### A case study on Dutch

#### Abstract

So-called linking elements may actually fulfil two very different roles: they may be phonotactic elements or they may realize a morphosyntactic head. I present a case study of linking elements in Dutch which do fulfil a morphosyntactic role. I note first that when 'linking elements' are of this type, the linguist should be aware of the fact that they may be null morphemes. I show that, indeed, Dutch may have such a null affix in its inventory. I then argue that the Dutch linking elements are in fact noun class markers and as such, I derive the fact that two of them are identical to plural marking through cumulative exponence. I point out that these noun classes co-exist with gender marking and may cause so-called erroneous reference. The analysis shows that a careful study of such elements provides us with new insights into two domains simultaneously. Firstly, we learn more about the morphosyntax of the compound's left-hand part, the structure of the compound itself and the inventory of compounding types in the language. The importance of the left-hand part cannot be underestimated in my opinion. It is a unique domain: it shows how a category behaves functionally in the absence of functional material that guarantees reference, such as number and the D-layer. As such, we learn, secondly, something about the morphosyntax of a category. In this contribution it appears that the Dutch nominal domain has nominal classes which are less visible in the nominal group as they are syncretic with number marking. Take away number marking and they become visible.

#### 3.1 Introduction

The term linking element is generally used when a phoneme or a string of phonemes can be found between two parts of a compound. The term suggests that the role of this element is to link the parts of the compound. This can indeed be observed in the following example from Brazilian Portuguese (example and analysis taken from Nóbrega 2020):

- (1) cervej-o-chato  
beer-LE-snob  
'beer snob'

The linking element -o- is there for phonotactic reasons. It connects a root that ends in a consonant and one that begins with one. It is absent in other phonological contexts, as can be seen in *hidrelectric* 'hydroelectric'. Note that, in (1) the vowel -o- does not belong to the left-hand noun, which would surface as *cerveja* 'beer'. The -o- is thus to be analyzed as a phonological element of the compounding structure, not as part of the compound's left-hand part. For good measure, Nóbrega (2014:211) analyses it as a dissociated morpheme. In other words, it would be a morpheme that is

inserted after syntax proper, but it would reflect syntactic structure, without realizing syntactic structure (see also Embick 1997). It is thus more than just an epenthetic vowel according to him.<sup>55</sup>

However, so-called linking elements may fulfil a different role than improving the phonotactics of a compound. They may realize a head in the functional sequence of the left hand's part morphosyntactic structure. The aim of this article is to show that this is indeed the case for the Dutch so-called linking elements *-s* and *-en*. I will argue that they realize class marking. For such affixes the term linking element is a misnomer. Their role is not to link two parts of a compound. Their function is to realize syntactic structure of the left-hand part of the compound.

More generally, the aim of this article is to draw attention to the fact that the term 'linking element' may refer to two very different linguistic elements: there are phonemes that improve the phonotactics of the compound and there are affixes that realize syntactic structure that belongs to the left-hand part of the compound. By presenting a case study, I aim to define criteria to approach the distinction.

This distinction corresponds to two types of primary compounds I have set apart in De Belder (2017). In De Belder (2017) I argued that Dutch has two types of primary compounds: those of which the non-head is a bare root (root primary compounds) and those of which the non-head bears nominal structure (nominal primary compounds). In that article I presented a discussion of the first type. The second type will be presented in this article. The root primary compounds discussed in De Belder (2017) never show any type of linking element in Dutch. The nominal primary compounds presented here show the so-called linking elements that in fact realize a piece of nominal functional structure.

As recently pointed out by Nóbrega (2020) the absence of a linking element is not a necessity for root primary compounds cross-linguistically. If it can be shown that a linking element is there for purely phonotactic reasons, and thus does not realize syntactic structure (as is consistent with his analysis of the *-o-* as a dissociated morpheme in (1)), it may still be possible that what precedes the linking element is a bare root, as in (1). In contrast, if it can be shown that the so-called linking element in fact realizes syntactic structure that belongs to the non-head, the non-head cannot be a bare root by definition, as it is marked with functional structure. Hence, in order to understand the typology of non-heads in compounds cross-linguistically, it is crucial to appreciate the role the 'linking element' fulfils in the compound.

A further example from Dutch can be found in De Belder & Van Koppen (2016). They discuss Dutch compounds of which the non-head contains an adjective. This may occur in primary compounds, when the non-head is an adjective or in more complex compounds, of which the non-head is a partial nominal group containing an adjective. Both types of compounds may show a schwa. De Belder and Van Koppen (2016) argue that this schwa is sometimes genuine adjectival inflection, which indeed, may be realized by a schwa in Dutch, as in the following example (2a):

- |     |    |                             |    |     |           |                     |
|-----|----|-----------------------------|----|-----|-----------|---------------------|
| (2) | a. | kaal-e-kat-en-adoptie       | b. | de  | kaal-e    | katten              |
|     |    | bald-ə-kat-PL-adoption      |    | the | bald-INFL | cats                |
|     |    | 'adoption of hairless cats' |    |     |           | 'the hairless cats' |

In (2a) the non-head of the compound is the partial nominal group *kale katten* 'hairless cats'. The schwa of *kale* 'bald' is indeed genuine adjectival inflection, as one would attest in a DP, see (2b).

---

<sup>55</sup> Eulàlia Bonet (p.c.) points out that an epenthetic vowel would surface as a schwa.

For other compounds that contain an adjective in the non-head, they argue that the schwa is merely there for phonotactic reasons. More specifically, they show that if the compound's head is a neuter noun, the schwa is unexpected as an inflectional marker on the adjective due to the details of Dutch adjectival inflection that should not concern us here. However, sometimes the schwa is present even though the head is a neuter noun. Yet, it is then always the case that the final consonant of the adjective is a dental or a lateral consonant. As a result, contrasting pairs such as the following ones can be found in Dutch. Example (3) shows compounds with a neuter noun as the head (the noun *jaar* 'year' is a neuter noun) and adjectives in the non-head position that differ phonologically: *oud* 'old' ends in a dental, whereas *nieuw* 'new' does not:

- |           |  |    |                                      |
|-----------|--|----|--------------------------------------|
| (3)    a. | oud-e-jaar<br>old-ə-year<br>'New Year's Eve' | b. | nieuw-jaar<br>new-year<br>'New Year' |
|-----------|--|----|--------------------------------------|

For the compounds where the schwa genuinely seems to functions as adjectival inflection, they proposed that there is adjectival functional structure present in the compound's non-head. The compounds that merely showed a phonotactic schwa were analyzed as root primary compounds. As such, they can be seen as an instance of compounds with linking elements that are there for phonotactic reasons, with the non-head being a bare root, as in Nóbrega's (2020) Brazilian Portuguese example above. The discussion illustrates the importance of a correct identification of the linking element as a syntactic head or a phonotactic phenomenon when analyzing specific compounds. A correct analysis of the structure of the compounds and the nature of a compounds' non-heads depends on a precise analysis of the role and the nature of the compounds' linking elements.

As pointed out above, this contribution presents one half of the story. I will show which characteristics can be found for linking elements that do fulfil a syntactic role. Readers who are interested in the other half of the story are referred to a discussion of root primary compounds in De Belder (2017), the discussion of adjectival non-heads in De Belder and Van Koppen (2016) and the discussion of linking elements in Brazilian Portuguese and Greek in Nóbrega (2020) and references therein.

This article is structured as follows. The next section presents the main data. Section 3 discusses the selection of the linking element. Section 4 argues that there may be a null affix in Dutch that fulfils the role of a linking element. Section 5 discusses the relation between linking elements and plural markers in Dutch: there seems to be a connection, but one cannot conclude that linking elements are plural markers. In section 6 it is argued that linking elements are class markers in Dutch and the analysis is fleshed out technically in section 7. Section 8 comments on the co-existence of noun classes and gender in Dutch. Section 9 concludes.

### 3.2 Dutch linking elements: main data

The non-head of a Dutch primary compound may be followed by either an *-s* or *-en*<sup>56</sup>, as in (4). These intervening morphemes have been called linking morphemes (Don 2009).<sup>57</sup>

- (4)    a. boek-en-kast  
            book-EN-case  
            'book case'  
  
            b. zon-en-klaar  
            sun-EN-bright  
            'very bright'  
  
            c. ezel-s-dracht  
            donkey-s-pregnancy  
            'long pregnancy'  
  
            d. hemel-s-blauw  
            sky-s-blue  
            'blue as the sky'

Previous studies have suggested that these linking elements are a piece of nominal inflection. This assumption stands to reason as it only follows nominal non-heads, as can be deduced from (5).<sup>58</sup>

- (5)    a. speur-hond  
            track-dog  
            'tracking dog'  
  
            b. drie-luik  
            three-panel  
            'triptych'  
  
            c. snel-trein  
            fast-train  
            'high-speed train'  
  
            d. achter-grond  
            back-ground  
            'background'

The debate has centered around the question which piece of nominal inflection is realized by the linking morpheme. The most dominant claims are that it is a remnant of case inflection or that it realizes plural marking (Neijt and Schreuder 2009) (and see Schreuder, Neijt, Van der Weide and Baayen 1998 for an overview of previous approaches). However, these proposals are unsatisfactory. The analysis which proposes that it etymologically goes back to case marking may be insightful (see Van Tiel, Rem and Neijt 2011 for a critical discussion), but it fails to provide an analysis of its present function. Given that present-day Dutch does not have overt case marking and given that the distribution of linking morphemes does not reflect the distributional patterns of case, the etymology does not lead us to an analysis of the facts in contemporary Dutch. The claim that they realize plural marking has much more plausibility as the linking elements are identical to Dutch' main plural markers. Yet, such an analysis faces empirical problems, as will become clear in section 5.

In this paper I provide an alternative proposal. I argue that linking morphemes realize noun class marking. I derive the close similarity between linking morphemes and plural markers through cumulative exponence. I propose that Dutch plural marking should be understood as an extension of noun class marking (cf. Harbour 2008 on Kiowa). Both projections, noun class and plurality, can

<sup>56</sup> As mentioned in the introduction, there is also -e- as in *oud-e-jaar* (old-e-year) ‘New Year’s Eve’ and *wit-e-brood* (white-e-bread) ‘white bread’, which is truly phonotactic and not under discussion in the present contribution. The string -er, as in *kind-er-wagen* (child-er-car) ‘pram’, belongs to an allomorph *kinder* from *kind* ‘child’ (see De Belder 2020).

<sup>57</sup> It will become clear the term 'linking morpheme' is a misnomer from a theoretical point of view as its purpose is not to link anything.

<sup>58</sup> See footnote 1 in which it is pointed out that the -e- in, for example, *oudejaar* is a phonotactic linking element and thus does not pattern with these examples.

be realized simultaneously by means of a single vocabulary item. It follows that linking morphemes and plural marking are highly similar, as they are realized by means of the same vocabulary item. Yet, they can be distinguished as well as they are two separate syntactic projections. I will further argue that noun class marking can be realized by means of a null vocabulary item, i.e. that the linking morpheme can be null. This again emphasizes the syntactic nature of these affixes, rather than a phonotactic nature. Finally, I point out that gender co-exists alongside number marking in Dutch. Such a view has been presented before for Romance languages (cf. Harris 1991 on Spanish), but it has been denied for Germanic languages (see, for example, Delfitto, Fábregas and Melloni 2008).

Given that the term linking morpheme is a misnomer both descriptively and theoretically for these facts, I will avoid using it much. I will call the morpheme the NCM, which can be interpreted as short for ‘nominal compound marker’, reflecting its descriptive properties or for ‘noun class marker’, matching the theoretical proposal in this contribution.

### 3.3 The selection of the NCM

In this section I will present some basic facts on Dutch NCMs. I will show that they are selected by the non-head in a regular fashion. As we have seen above, Dutch has two overt NCMs, viz. *-s* and *-en*, as in (6).

- |   |   |
|---|---|
| (6) a. peer-en-boom<br>pear-NCM-tree<br>'pear tree' | b. varken-s-hok<br>pig-NCM-pen<br>'pig's pen' |
|---|---|

The NCM *-en* may be pronounced as ə, ən or as a syllabic n, depending on the dialect and the phonological context of the word (Hanssen 2012). Below we will see there is perhaps yet a third NCM, which is a null morpheme.

The NCM is selected by the compound’s non-head. For example, *kat* ‘cat’ selects *-en*, as in (7), whereas *koning* ‘king’ selects *-s*, as in (8).

- |   |  |  |  |
|---|--|--|--|
| (7) a. kat-en-luik<br>cat-NCM-shutter<br>'cat flap'         | b. kat-en-voer<br>cat-NCM-food<br>'cat food'         | c. kat-en-staart<br>cat-NCM-tail<br>'cat tail'         | d. kat-en-bak<br>cat-NCM-box<br>'cat litter box' |
|   |  |  |  |
| (8) a. koning-s-kind<br>king-NCM-child<br>'child of a king' | b. koning-s-kroon<br>king-NCM-crown<br>'royal crown' | c. koning-s-troon<br>king-NCM-throne<br>'royal throne' |  |
|   |  |  |  |
| d. koning-s-paar<br>king-NCM-couple<br>'royal couple'       |  |  |  |

If two non-heads are coordinated, they still select their own NCM, as in (9) (Botha 1968). This indicates that the head of the compound does not influence the selection of NCM.

- (9) dame-s- en heer- en jassen  
 lady-NCM- and gentleman- NCM- coats  
 ‘coats for ladies and gentlemen’

The selection of the NCMs is much more regular than commonly assumed. Admittedly, their distribution seems somewhat irregular in Standard Dutch. For example, (10)a and (10)b exist alongside one another (see the dictionary *Van Dale*). They both contain the non-head *stad* ‘city’, yet (10)a selects -s, whereas (10)b selects Ø.



Yet, native speakers share intuitions on the selection of NCMs. If one asks a native speaker to form a compound with *stad* 'city' as the non-head and *hond* 'dog' as the head, the result, which is shown in (11), can be predicted.

- (11) stad-s-hond  
city-NCM-dog  
'dog suited for/living in a city'

This point was confirmed in a questionnaire which was filled out by 689 native speakers of Dutch. I asked native speakers to indicate their preferred NCM in newly formed compounds. More than 90% of the speakers agree in most cases. For example, 95% of the speakers agree the neologism in (12)a is better than its competitors in (12)b and (12)c.



Admittedly, for some non-heads there is variation. For example, 49% of the native speakers prefer (10)a, but 56% of the speakers prefer (13)b.<sup>59</sup> None of them like (13)c.

- (13) a. pater-Ø-korting  
priest-NCM-reduction  
'reduction for priests'      b. pater-s-korting  
priest-NCM-reduction  
'reduction for priests'      c. pater-en-korting  
priest-NCM-reduction  
'reduction for priests'

I will come back to this issue in section 6.3. I will point out that even though one cannot predict a native speaker's choice between *-s* and  $\emptyset$  in (13), it still can be predicted that exactly this non-head will trigger variation and that *-en* will not be selected. In sum, it is possible to draw descriptive generalizations on the distribution of NCMs.

<sup>59</sup> The sum is more than 100% because the informants were able to choose both forms in case they had no preference. Some speakers thus accepted both forms equally.

Observe that neologisms which have been listed only recently in Dutch dictionaries are completely regular. It can be checked in dictionaries that newly created compounds with *stad* ‘city’ as their non-head invariably select -s as a NCM, as in (14).

- |                              |                              |
|------------------------------|------------------------------|
| (14) a. <i>stad-s-eiland</i> | b. <i>stad-s-cowboy</i>      |
| city-NCM-isle                | city-NCM-cowboy              |
| ‘urban isle’ <sup>60</sup>   | ‘urban cowboy’ <sup>61</sup> |

I will therefore assume that the NCM which appears in neologisms is the regular one. The fact that neologisms show a regular selection of NCMs in contrast with older compounds dovetails with the fact that compounds with NCMs are only attested relatively recently in Dutch. They only appear since the sixteenth century and they are only attested frequently since the twentieth century (van Tiel, Rem and Neijt 2011). This indicates that the irregular forms as in (10)b are but remnants of older word-formation types. As such, they do not *per se* indicate that the selection of NCMs is irregular. It merely shows that the lexicon contains words which originated in different periods and which are preserved in their original form in the lexicon. I therefore propose to mainly ignore lexicalized compounds when drawing generalizations and to focus on productively formed new compounds in the discussion (cf. Bauer 1978:78 who makes the same point).

### 3.4 The NCM can be null

A crucial prediction that sets apart affixes that realize syntactic heads from those that are there for phonotactic reasons (see introduction) is that the syntactic affixes may be null affixes, whereas the phonotactic linking elements may not. Phonotactic null elements, in the sense of epenthetic elements, cannot exist by definition.<sup>62</sup> Yet, if NCMs realize nominal inflection, one may certainly hypothesize they may be null. After all, it is common for inflection in Dutch to be realized by means of a null vocabulary item. A In this section I will argue that there is indeed a null NCM alongside the overt ones -s and -en.

Compounds with a nominal non-head followed by a NCM are not the only type of primary compounding in Dutch. As mentioned in the introduction, De Belder (2017) argues there are two types of primary compounds in Dutch. The first type is the one under discussion in the present article. It is characterized by a nominal non-head and the presence of an NCM (these compounds are called nominal primary compounds in De Belder 2017). The second type contains a bare root as its non-head which is never followed by any overt affixes (these compounds are called root primary compounds in De Belder 2017). Examples of root primary compounds can be found in (15).

---

<sup>60</sup> A *stadseiland* is an artificial island which is created for urban development. The lemma appeared in the dictionary Van Dale in 2007 (Dikke Van Dale Online, no authors (2020)).

<sup>61</sup> The lemma *stadscowboy* appeared in the dictionary *Van Dale* dictionary in 2009.

<sup>62</sup> To be entirely clear: if the only ground for the presence of an exponent is a phonological one (i.e. if the exponent would optimize the phonological surface representation of the compound), it would not make any sense to get a zero exponent; it would defeat its own purpose. Eulàlia Bonet points out that languages may have exponents the overt presence of which can be motivated by an interplay of phonological and syntactic factors (Bonet 2018, see also Zygis 2010 and De Belder to appear). The present reasoning does not apply to such exponents: their overt or null realization would depend on syntactic factors as well. I am not aware of studies of such exponents in compounds, but they are certainly a hypothetical possibility.

- (15) a. speur-hond  
track-dog  
'tracking dog'
- b. drie-luik  
three-panel  
'triptych'
- c. snel-trein  
fast-train  
'high-speed train'
- d. achter-grond  
back-ground  
'background'

The fact that the non-head may be a bare root results in a methodological problem. How to set apart those compounds which have a bare root not followed by anything as their non-head (root primary compounds) and those compounds which contain a null NCM (nominal primary compounds)? At the surface, they are both expected to look the same.

A puzzling data set in this respect are compounds of which the non-heads refer to mass concepts. Such compounds never show an overt NCM.

- (16) a. kaas-boer  
cheese-farmer  
'cheese farmer'
- b. zand-grond  
sand-soil  
'sandy soil'
- c. vuur-steen  
fire-stone  
'flint'
- d. siroop-fles  
syrup-bottle  
'syrup bottle'
- e. klei-grond  
clay-soil  
'clay soil'
- f. wol-draad  
wool-yarn  
'wool-yarn'

We may now ask the question whether such non-heads systematically occur in root primary compounds or whether they systematically select a null NCM. In Standard Dutch, it is, to the best of my knowledge, impossible to find empirical support for any of these two hypotheses.

West-Flemish Dutch data shed light on the issue. Nouns are marked for gender in Dutch. Yet, whereas in Standard Dutch the gender of the noun is only noticeable via agreement on the adjective and the determiner, West-Flemish feminine nouns are also marked via an obligatory schwa ending on the noun (Haegeman 1998), as in (17).<sup>63</sup>

- (17) e school-e  
a school-F  
'a school'

As in Standard Dutch, mass concepts are not followed by an overt marker in West-Flemish when they are the non-head in a primary compound:

- (18) a. de koas  
the cheese  
'the cheese'
- b. het zand  
the sand  
'the sand'
- c. het vier  
the fire  
'the fire'

---

<sup>63</sup> The West-Flemish Dutch examples in this article are taken from Blankenberge Dutch. I would like to thank Katlijn Van Audenaerde and Monica Roose for data and judgments.

- (19) a. koas-boer                    b. zand-groend                    c. vier-stêen  
       cheese-farmer                sand-soil                        fire-stone  
       'cheese farmer'              'sandy soil'                  'flint'

However, crucially, this fact only holds for masculine and neuter mass nouns in this dialect, as can be seen in (60). (18) shows that the non-heads are indeed not feminine, as they do not select the schwa marker. In (18), the nouns preceded by the definite article *de* are masculine, those preceded by the definite article *het* are neuter nouns. If the root which refers to a mass concept is feminine, as in (20), its schwa ending is retained within the compound, as in (21).<sup>64</sup>



The fact that gender marking is present in these examples indicates that the root is followed by nominal inflection in general. Hence, the non-head in these compounds is not a bare root, but a root under nominal inflection. If we do not want to multiply the assumed types of compounding in Dutch beyond necessity, it stands to reason that West-Flemish compounds containing a feminine mass noun followed by a gender marker as their non-head are nominal primary compounds. More generally, there is evidence in Dutch that primary compounds with a mass noun as their non-head can be nominal primary compounds. I extend this conclusion to Standard Dutch and I assume that compounds with a root referring to a mass concept may be nominal primary compounds as well. They then do contain nominal markers, albeit covertly. I therefore postulate the presence of a null NCM for the data in (16).

For good measure, note that the feminine marker is not obligatorily present in West-Flemish compounds. Just as Standard Dutch (and Germanic languages in general, see De Belder 2017), West-Flemish Dutch certainly has root primary compounds which may contain roots that may occur as feminine nouns. In such compounds, the feminine marker is absent in West-Flemish Dutch, as expected:

- (22) a. vrouw-mens                    b. stroat-kat-e                    c. school-gebouw  
    woman.FEM-human                street.FEM-cat-FEM                school.FEM-building  
    ‘woman (pejorative)’                ‘street cat’                        ‘school building’

d. kliet-soort-e  
    clay.FEM-kind-FEM  
    ‘kind of clay’

<sup>64</sup> Note that, as expected, the schwa is not present in Standard Dutch examples, which lack an overt gender marking on the noun (e.g. *siroop-fles* 'syrup bottle' and *klei-grond* 'clay soil').

Loanwords behave similarly as mass nouns. There is no overt NCM in Standard Dutch when the compound's non-head is a loanword:

- |                      |                   |              |
|----------------------|-------------------|--------------|
| (23) a. Ford-fabriek | b. dollar-koers   | c. café-baas |
| Ford-factory         | dollar-currency   | pub-holder   |
| 'Ford factory'       | 'dollar currency' | 'pub holder' |

Haegeman (1998) observes that if a loanword happens to end in a schwa, it will be treated as a feminine noun in West-Flemish:

- |                  |                   |
|------------------|-------------------|
| (24) a. e lift-e | b. e disket-e     |
| a.F elevator-F   | a.F floppy.disk-F |
| 'an elevator'    | 'a floppy disk'   |

Now, my informants often allow both nominal primary compounds and root primary compounds in West-Flemish:

- |                            |                          |                  |
|----------------------------|--------------------------|------------------|
| (25) a. lift-(e)-telefong  | b. disket-(e)-hoesje     | c. col-e-fabriek |
| elevator-F-telephone       | floppy.disk-F-case       | glue-F-factory   |
| 'telephone in an elevator' | 'case for a floppy disk' | 'glue factory'   |

This suggests that the Standard Dutch data in (16) probably underlyingly show variation between root primary compounding and nominal primary compounding as well. When they are nominal primary compounds, the NCM is a null affix. Of course, we are extending here a logic established on the basis of dialectal data to Standard Dutch data. Within Standard Dutch, the data themselves do not reveal anything relating to this matter.

I would like to note that there are classes of nouns that do not show this type of variation. Nouns that refer to body parts and kinship terms (a pair of semantics fields that is known as a natural class of referring to inalienable possession) never show an NCM in Standard Dutch:

- |                       |                        |                 |                |
|-----------------------|------------------------|-----------------|----------------|
| (26) a. hoofd-pijn    | b. arm-band            | c. lever-ziekte | d. huid-ziekte |
| head-ache             | arm-band               | liver-disease   | skin-disease   |
| 'head ache'           | 'bracelet'             | 'liver disease' | 'skin disease' |
|                       |                        |                 |                |
| (27) a. broeder-moord | b. dochter-knoop       | c. vader-land   |                |
| brother-murder        | daughter-node          | father-land     |                |
| 'fratricide'          | 'daughter node'        | 'fatherland'    |                |
|                       |                        |                 |                |
| d. oom-zegger         | e. zoon-wens           | f. moeder-taal  |                |
| uncle-adresser        | son-wish               | mother-language |                |
| 'nephew/niece'        | 'desire to have a son' | 'mother tongue' |                |

Yet, they never show feminine markers in West-Flemish Dutch compounds either:

- (28) a. e nek-e  
       a.F neck-F  
       'a neck'
- b. nek-(\*)e)-klachtn  
       neck-(\*)F)-complaints  
       'neck complaints'

- (29) a. e moedr-e  
       a.F mother-F  
       'a mother'
- b. moeder-(\*)e)-melk  
       mother-(\*)F)-milk  
       'breast milk'

It seems that body parts and kinship terms are limited to primary root compounds. For good measure, I would like to note that proper names systematically resist overt NCMs as well in Standard Dutch:

- (30) a. Nijl-delta  
       Nile-delta  
       Nile delta'
- b. Vitra-stoel  
       Vitra-chair  
       'Vitra chair'
- c. Mozart-kenner  
       Mozart-specialist  
       'Mozart specialist'

Whether proper names pattern with mass nouns and loanwords or with kinship terms and body parts is unclear to me.

In sum, this section aimed to point out that what is called ‘a linking element’ may be there for syntactic reasons and, as soon as one accepts that, there may be a null affix that fulfils this function. I aimed to point out that it is reasonable to assume that Dutch indeed has such a null affix. The next section discusses which piece of nominal inflection is realized by the linking element. I first argue they are not plural markers. In a later section I will propose they realize noun class marking.

### 3.5 NCMs and plural markers

In this section I discuss the close relation between NCMs and plural markers: they are similar formally and distributionally and psycholinguistic experiments show that native speakers associate NCMs with plural markers. Yet, it can be shown that NCM are not plural markers.

Let us first consider the formal relation between NCMs and plural markers. Recall that the two overt NCMs in Dutch are *-s* and *-en*. Dutch’ main<sup>65</sup> plural markers are formally identical, they are *-s* and *-en* as well, as is illustrated in (31).

- (31) a. twee ezel-s  
       two donkey-PL  
       'two donkeys'
- b. twee kat-en  
       two cat-PL  
       'two cats'

Hanssen (2012) has shown that this formal identity is absolute in all Dutch dialects. The NCM *-en* may be pronounced as *ə*, *ən* or as a syllabic *n*, depending on the dialect and the phonological context of the word. It now appears that regardless how the NCM *-en* is pronounced in a given dialect, its

---

<sup>65</sup> I am ignoring plural markers on loan words, such as *-a* in *musea* ‘museums’.

pronunciation will be identical to the pronunciation of the one of the plural marker *-en*.<sup>66</sup> The set of NCMs and the set of plural markers are thus formally identical across dialects.

NCMs and plural markers show a certain relation in their distribution. Mattens (1984) observed the generalization in (32).

(32) The *s*-plural generalization

If a root takes *-s* as its plural marker, it will not take *-en* as its NCM in the same idiolect.

For example, from the fact that *ezel* ‘donkey’ takes *-s* as its plural marker, as is shown in (33), we predict correctly that (34) is excluded.

(33) twee ezel-s

two donkey-PL

‘two donkeys’

(34)\* *ezel-en-dracht*

donkey-LP-pregnancy

There is thus a certain distributional relation between NCMs and plural markers. I come back to this relation in more detail below.

Finally, there is a psychological relation between NCMs and plural markers. Psycholinguistic experiments show that native speakers associate the NCM with plurality, even if such an association is contextually implausible, as in *zwaluw-en-ei* (swallow-NCM-egg, the egg comes from a single bird) (Schreuder, Neijt, Van der Weide and Baayen 1998, Neijt, Baayen en Schreuder 2006, Neijt and Schreuder 2009). In sum, there is a clear link between NCMs and plural markers. They are formally identical, they are related in their distribution and native speakers associate them with one another.

Unsurprisingly, it has been proposed that NCMs *are* plural markers (see Neijt and Schreuder 2009 for an example of such an analysis). Yet, such an analysis faces four empirical problems. Firstly, it has been noted that semantic plurality is not a predictor of the selection of a NCM (Booij 2002). The non-head of the compounds in (35) are interpreted as singular. Yet, they do select a NCM. If the NCM were a plural marker, it is not clear how the meaning of the examples in (35) is derived.<sup>67</sup>

- (35) a. hond-en-drol                    b. vrouw-en-lijk                    c. rijtje-s-huis  
      dog-NCM-turd                      woman-NCM-corps              row-NCM-house  
      ‘turd of a dog’                      ‘corpse of a woman’              ‘row house’

The second problem is the fact that the NCM may arguably be null (see section 4), whereas there is no null plural marker in Dutch. In other words, there is at least one NCM which cannot be analyzed as a plural marker.

The third empirical problem is the fact that the NCM which is selected by a specific root does not necessarily match the plural marker which is selected by this root. For example, the root *dorp* selects *-s* as its NCM, as in (36)a, yet its plural marker is *-en*, as in (36)b.

<sup>66</sup> In contrast, the pronunciation between the plural marker *-en* and the plural agreement marker *-en* on the verb may differ.

<sup>67</sup> Admittedly, the strength of this argument depends on the precise semantic analysis of plurality.

- (36) a. dorp-s-café  
          village-NCM-pub  
          'pub in a village'
- b. twee dorp-en  
          two village-PL  
          'two villages'

Similarly, roots selecting a null NCM, selects -s or -en as their plural marker:

- (37) a. koffie-∅-kop  
          coffee-NCM-mug  
          'coffee mug'
- b. twee koffie-s  
          two coffee-PL  
          'two coffees'
- (38) a. kaas-∅-boer  
          cheese-NCM-farmer  
          'cheese farmer'
- b. twee kaaz-en  
          two cheese-PL  
          'two kinds of cheese'

Fourthly, we have seen above that there is a distributional relation between NCMs and plural markers: if a root takes -s as its plural marker, it will not take -en as its NCM in the same idiolect. This generalization certainly points to a connection between the two affixes, but it does not follow that they are one and the same affix. Furthermore, from the plural marker -en no prediction can be derived with regards to the NCM. In other words, a root which selects -en as its plural marker may select either NCM. These facts are summarized in the table below. They will play a central role in the analysis in section 7.

(39)

	Plural Marker	Noun Class Marker
s-plurals	s	∅
	s	s
en-plurals	en	∅
	en	s
	en	en

Under the assumption that NCMs are plural markers, the lack of a one-to-one mapping in the distribution cannot be captured.

The relation between NCMs and plural markers in Dutch is puzzling. They are clearly related, yet they cannot be equated. How, then, should we understand the close relation between the NCMs in Dutch and plural marking? An analysis of NCMs as noun class markers is particularly promising to answer this question. Cross-linguistically, it is extremely common for noun class

markers to be syncretic with plural markers (Kihm 2005:462). As such, noun class markers and plural markers may be closely related, yet not identical.

### 3.6 NCMs realize noun class marking

#### 3.6.1 Introduction

In this section I argue that NCMs are noun class markers in Dutch. Considering cross-linguistic properties of noun class markers (Harris 1991, Aikhenvald 2000, Kihm 2005, Harbour 2008), we know that if Dutch had them, they would be restricted to nouns, syncretic with plural markers, distributed according to their morphological and phonological and/or semantic properties and they would trigger agreement in the DP or clause. We have seen in the previous sections that NCMs and plural markers are indeed formally related. In this section I will show that Dutch NCMs show precisely those other properties as well.<sup>68</sup>

Noun class markers are of course part of the nominal domain. We have seen in section 2 that NCMs are indeed restricted to compounds with nominal non-heads. Non-nominal ones were restricted to root primary compounds and did not select an NCM. NCMs are thus a typical property of nominal non-heads. As such, they qualify for the first and most evident requirement of noun class marking. In the next sections, I will argue that they show more particular characteristics of noun class marking as well.

In the previous sections we have seen that the understood plurality of the compound's non-head is not a predictor of the selection of the NCM. One may wonder then what does predict its selection. Although a complete description of this matter transcends the ambitions of this article, I will point out that the choice of the NCM is determined by exactly those features which are known to define noun classes. These are the phonological, semantic and morphological properties of the root (Aikhenvald 2000).

#### 3.6.2 Phonological criteria

Phonological criteria play a role in the selection of the NCM. Note first that the distribution of plural marker *-s* is partially determined by phonological criteria. Some roots trigger the plural marker *-s* for phonological reasons. I will refer to these roots as *-s*-plurals: Roots which take *-s* as their plural marker for phonological reasons (in a given dialect/idiolect/register) are called *s*-plurals (for that dialect/idiolect/register).

*S*-plurals include two main groups.<sup>69</sup> The first group are roots which end in a schwa followed by a sonorant. Examples are given in (40).

- (40) a. bodem-s      b. vogel-s      c. toren-s      d. letter-s  
bottom-PL      bird-PL      tower-PL      letter-PL  
'bottoms'      'birds'      'towers'      'letters'

<sup>68</sup> Harris (1991) notes that class markers are not exclusive to nouns, they occur on adverbs as well. Interestingly, Dutch adverbs often end in an *-s*, see Corver (2017) for discussion.

<sup>69</sup> An exhaustive discussion of the relation between the phonological make-up of the root and the selection of the plural marker would take us too far afield. For a detailed discussion see Kooij & van Oostendorp (2003).

The second group are roots which end in an open vowel, which is not /i/ or /e/, as in (31).

- |                 |           |               |
|-----------------|-----------|---------------|
| (41) a. piano-s | b. mama-s | c. tiramisu-s |
| piano-PL        | mommy-PL  | tiramisu-PL   |
| 'pianos'        | 'mommies' | 'tiramisus'   |

Now recall from section 5 that there is a distributional link between the selection of the plural marker and the NCM. The relevant generalization is repeated in (42).

- (42) The *s*-plural generalization

If a root takes -*s* as its plural marker, it will not take -*en* as its NCM in the same idiolect.

It follows that roots which end in a schwa followed by a sonorant or in an open vowel, which is not /i/ or /e/ will select either -∅ - or -*s*- as their NCM, but not -*en*- As such, the selection of NCMs is partially determined phonologically.

### 3.6.3 Conceptual criteria

Conceptual properties of the non-head influence the selection of the NCM. These properties include mainly masshood, animacy and concreteness (see Mattens 1984). There are other minor relevant factors such as granularity and edibility. These and comparable properties support the hypothesis that NCMs are noun class markers as they are textbook examples of the relevant criteria which determine noun classes cross-linguistically (Aikhenvald 2000).<sup>70</sup>

A study on lexical statistics (Krott et al. 2009) has shown that Dutch non-heads referring to animate and concrete concepts tend to get the NCM -*en*, whereas abstract roots tend to select -*s* and inanimate ones favor ∅. Telling though these results may be, they focus on coarse distinctions. Below I present some more fine-grained observations.

As we have seen above, non-heads which refer to mass concepts select the NCM ∅, regardless whether they are *s*-plurals or *en*-plurals (see also Mattens 1984). More examples are given below. (43) shows compounds with *s*-plurals as their non-heads, (44) shows compounds with *en*-plurals.

- |                      |                  |                  |
|----------------------|------------------|------------------|
| (43) a. ijzer-∅-erts | b. koper-∅-mijn  | c. water-∅-bron  |
| iron-NCM-ore         | copper-NCM-mine  | water-NCM-source |
| 'iron ore'           | 'copper mine'    | 'water source'   |
| d. gember-∅-koek     | e. amber-∅-steen | f. balsem-∅-geur |
| ginger-NCM-cookie    | amber-NCM-stone  | balsam-NCM-scent |
| 'ginger cookie'      | 'amber stone'    | 'balsam scent'   |
- 
- |                     |                 |                  |
|---------------------|-----------------|------------------|
| (44) a. wijn-∅-glas | b. melk-∅-fles  | c. siroop-∅-fles |
| wine-NCM-glass      | milk-NCM-bottle | syrup-NCM-bottle |
| 'wine glass'        | 'milk bottle'   | 'syrup bottle'   |

---

<sup>70</sup> These properties are unified by the property of influencing the individuation of the referent (see Grimm 2012).

d. kaas-∅-boer	e. honing-∅-wijn	f. zand-∅-strand
cheese-NCM-farmer	honey-NCM-wine	sand-NCM-beach
'cheese farmer'	'honey wine'	'sandy beach'
g. vuur-∅-steen	h. ijs-∅-gletsjer	i. licht-∅-straal
fire-NCM-stone	ice-NCM-glacier	light-NCM-beam
'flint'	'ice glacier'	'light beam'

There is one class of exceptions. If the root is an *en*-plural which refers to an edible grain, it may select *-en*, even though it may be considered to be a mass concept. Examples are given in (45).

- (45)      a. gerst-en-bier                          b. gort-en-pap                          c. spelt-en-meel  
               barley-NCM-beer                         pearl.barley-NCM-porridge                 spelt-NCM-flour  
               'barley beer'                                'porridge of pearl barley'                    'spelt flour'

These examples may seem particular exceptions, yet an interplay between masshood and granularity is not unique to Dutch noun classes. In Kiowa, for example, non-granular mass nouns belong to one class, whereas granular mass nouns belong to another one (Harbour 2008:50). Hence, *áápenhaa* 'honey, syrup' belongs to a different class than *éiyóguei* 'rice' in Kiowa. Admittedly, the Dutch data are not fully parallel to the data in Kiowa. In Dutch it does not suffice to be granular for a mass noun to belong to the exceptional class. The granular mass concept needs to be edible and vegetable (i.e. a grain) as well. Yet, the underlying similarity is striking.

For Dutch non-heads which end in a schwa plus a sonorant animacy is relevant. Mattens (1984) observed that inanimate roots of this type select ∅, whereas animate ones select -s. I have checked these conclusions in a questionnaire, to which 689 speakers of Dutch responded. It appears that his generalizations should be adjusted slightly. Roots which end in a schwa plus a sonorant and which refer to inanimate concepts indeed select ∅ in neologisms,<sup>71</sup> as illustrated in (46). However, roots referring to animals select ∅ as well, as in the examples in (47).<sup>72</sup>

- (46)      a. bodem-∅-hut                          b. deken-∅-kast                          c. kabel-∅-web  
               bottom-NCM-cabin                        blanket-NCM-closet                        cable-NCM-web  
               'bottom cabin'<sup>73</sup>                        'closet for blankets'                    'web consisting of cables'

---

<sup>71</sup> I was not interested in NCMs which occur in lexicalized compounds. I tested newly formed compounds (see section 3 for a justification of this approach). Given that I invented neologisms for the test, it should not come as a surprise that some examples in (46)-(49) are slightly unconventional meaningwise. In the questionnaire I made sure to provide the informants with a background context to render the neologism plausible from a pragmatic point of view.

<sup>72</sup> From the test it appeared that *varken* 'pig' exceptionally selects -s as its NCM. We may account for this observation if we suppose that *varken* is morphologically complex, with *vark* being the root and *-en* being a suffix. This hypothesis is supported by the fact that one may substandardly attest the diminutive *vark-ske* 'little pig', indicating that *vark* is a root. Below I point out that in morphologically complex non-heads the affix selects the NCM. I thus capture the fact that the selection of the NCM by *varken pig* is exceptional.

<sup>73</sup>In the test this neologism referred to a newly invented cabin in a zoo. The cabin had a transparent bottom which allows visitors to see insects living in the earth.

- (47) a. wezel-∅-tuin  
weasel-NCM-garden  
'garden for weasels'  
b. ezel-∅-taart  
donkey-NCM-pie  
'pie for donkeys'  
c. otter-∅-tuin  
otter-NCM-garden  
'garden for otters'

Roots which refer to human beings pattern differently. They allow variation between the NCM -s and ∅, as in (48) and (49).

- (48) a. dokter-s-babbel  
doctor-NCM-chat  
'chat with the doctor(s)'  
b. minister-s-lied  
minister-NCM-song  
'song for the minister'  
c. pater-s-korting  
priest-NCM-reduction  
'reduction for priests'
- (49) a. dokter-∅-babbel  
doctor-NCM-chat  
'chat with the doctor(s)'  
b. meester-∅-huisje  
teacher-NCM-house.DIM<sup>74</sup>  
'small house for teachers'  
c. pater-∅-korting  
priest-NCM-reduction  
'reduction for priests'

Although I fail to understand the variation between these two NCMs, it is striking that it can be predicted which class of roots is subject to it. They all end in a sonorant plus a schwa and they all refer to human beings. Despite the idiolectal variation a specific semantic field is singled out when it comes to the selection of NCMs. Again, a parallel can be drawn with languages with *bona fide* noun classes. Human beings form a separate class in, for example, Manjaku. In this language the noun *na-kiēj* 'thief' selects the noun class marker *na-* of class 1 in the singular, whereas the noun *u-ndali* select the class marker *u-* of class 3 in the singular (Kihm 2005:464).

Another example of a semantic field which patterns alike when it comes to the selection of NCMs are non-heads which refer to geographical units.<sup>75</sup> They select the NCM -s, even when they are all *en*-plurals.

- (50) a. dorp-s-café  
village-NCM-pub  
'pub in a village'  
b. stad-s-bus  
city-NCM-bus  
'city bus'  
c. staat-s-bank  
state-NCM-bank  
'state bank'  
d. land-s-verraad  
land-NCM-betrayal  
'treason'  
e. rijk-s-wacht  
state-NCM-guard  
'gendarmerie'

All non-heads in (50) are *en*-plurals, yet they all select s as their NCM. One might conjecture that the roots in (50) do not form a class and that the observation is due to happenstance. Yet, morphologically simplex *en*-plurals which select -s as their NCM are rather rare. As such, the examples in (50) genuinely pattern alike. This example of a unified selection of the NCM is again tied to the fact that they share a semantic property. The fact that locative nouns pattern alike when

---

<sup>74</sup> DIM = diminutive

<sup>75</sup> These are new observations, they were not included in Mattens (1984).

it comes to noun classes has been attested cross-linguistically as well. For example, KiVunjo-Chaga has a separate class for such nouns (Moshi 1995).

Similarly, morphologically simplex en-plurals referring to a geographic unit of water will select the null morpheme as well:

- (51) a. rivier-∅-bedding      b. zee-∅-visser      c. meer-∅-aal  
      river-NCM-bed            sea-NCM-fisherman     lake-NCM-eel  
      'river bed'                'sea fisherman'        'lake eel'  
  
      d. beek-∅-forel  
      creek-NCM-trout  
      'creek trout'

In sum, we have seen that the non-head's conceptual properties determine the selection of the NCM. Relevant factors are notions such as animacy, masshood, concreteness, granularity, edibility, geographical units,... In this respect, NCMs pattern with *bona fide* noun class markers cross-linguistically. I conclude that the observations support the hypothesis that NCMs are noun class markers.

### 3.6.4 Noun class agreement

Now, if Dutch nouns are marked for class, we may perhaps expect to notice the existence of noun classes outside of the domain of compounds. This is indeed the case. More specifically, there is an agreement pattern I would like to draw attention to.

Let me start with a small detour. Notice that Dutch has *bona fide* gender marking on nouns. Northern Dutch distinguishes between common and neuter gender, southern Dutch distinguishes between masculine, feminine and neuter gender. There is no correlation whatsoever between the classes we have observed above and gender marking in Dutch. For example, the group of nouns in (50) contain both common nouns (*stad, staat*) and neuter nouns (*dorp, land, rijk*). Gender is visible via agreement on definite articles, adjectival agreement and on indefinite articles in some dialects. Pronouns that refer back to a noun may also agree in gender with that noun in the more southern variants of Dutch:

- (52) De deur, **ze** staat open.  
      the door.F, she stands open  
      'As to the door, it is open.'

For some nouns, it has been observed that the reference may be 'erroneous'. In (53) a feminine possessive pronoun refers back to a neuter noun, in (54) a neuter pronoun refers back to a masculine noun.

- (53) % **Het<sub>NEUTER</sub>** volk<sub>NEUTER</sub> heeft haar<sub>FEM</sub> eigen mening. (Belgian Dutch)  
      the people has her own opinion  
      'The public has its own opinion.'

- (54) % Vis<sub>MASC</sub>, ik vind dat<sub>NEUTER</sub> lekker. (Belgian Dutch)  
 fish, I find it tasty  
 'Fish, I like it.'

An in-depth study (Audring 2009) reveals that the so-called erroneous reference is not random. The following morphologically simplex 'collective' nouns can be referred to with a feminine pronoun, even though they are neuter or masculine:

- (55) raad<sub>MASC</sub> 'council' bestuur<sub>NEUTER</sub> 'governing board'  
 staat<sub>MASC</sub> 'state' volk<sub>NEUTER</sub> 'public'  
 tijd<sub>MASC</sub> 'time' publiek<sub>NEUTER</sub> 'public'  
 arbeid<sub>MASC</sub> 'labor'

Mass DPs can always be referred to with a neuter pronoun, regardless of their gender, as in the example in (54), in which *fish* gets a mass reading.

I would like to point out that these erroneous references correlate with the classes I have proposed above. We have seen that mass readings constitute a class on their own, triggering a null NCM. The collective nouns all behave on a par as well. The table below shows that their gender can vary, but they can all select an 'erroneous' feminine referring pronoun, they all select the -s- as an NCM, as was shown in (50) and they all get the same plural marker *-en*, which differs from their plural marker, which is invariably *-en*.<sup>76</sup>

Table 1: the class of abstract collections

noun	gender	erroneous ref	Noun Class Marker	Plural Marker
raad 'council'	masc	fem	s	en
staat 'state'	masc	fem	s	en
tijd 'time'	masc	fem	s	en
arbeid 'labor'	masc	fem	s	unattested
bestuur 'board'	neuter	fem	s	en
volk 'public'	neuter	fem	s	en
publiek 'public'	neuter	fem	s	en

We now observe that even though the singular DP is not marked for class overtly by means of a marker, it does show a grammatical effect regulated by class by means of erroneous reference.

---

<sup>76</sup>The noun *arbeid* is quite rare in general and not commonly used in plural contexts. The plural form of *volk* is *volkeren*: it involves the allomorph *volker-*.

Structurally, it must be the case that the singular DP contains a class projection. It must further be the case that the compound's left-hand part contains this projection as well.

### 3.6.5 Morphological criteria

Before closing off this section, I would like to make a final remark for good measure. The morphological make-up of the non-head may influence the selection of the NCM as well. More specifically, the findings presented above are only relevant for non-heads which do not contain suffixes. Suffixes select their own NCMs. For example, *-heid* '-ity', as in (56), and the diminutive morpheme, as in (57) select *-s*.

- |  |  |  |
|--|--|--|
| (56) a. werkelyk-heid-s-besef<br>real-HEID-LP-sense<br>'sense of reality'      | b. schoon-heid-s-ideaal<br>beautiful-HEID-LP-ideal<br>'beauty ideal' | c. waar-heid-s-waarde<br>true-HEID-LP-value<br>'truth value' |
| (57) a. bloem-etje-s-behang<br>flower-DIM-LP-wallpaper<br>'flowered wallpaper' | b. rij-tje-s-huis<br>row-DIM-PL-house<br>'terraced house'            | c. kraaan-tje-s-water<br>tap-DIM-LP-water<br>'tap water'     |

It may be clear for now that the selection of the NCM may depend on a suffix. If NCMs are analyzed as noun class markers, it should not come as a surprise that their selection can be determined by affixes. This phenomenon has been attested in various languages (Aikhenvald 2000:25). I will not discuss this fact any further.

### 3.6.6 Conclusion

This section focused on the fact that Dutch NCMs show the prototypical features of noun class markers. They are restricted to nouns, they are syncretic with plural markers, they are selected on the basis of the phonological, conceptual and morphological properties of the noun and they show agreement.

Considering the non-head's phonological, conceptual or morphological properties or a combination thereof is the best way to predict the NCM it will select. Phonologically, Dutch nouns fall into two classes, i.e. *s*-plurals and *en*-plurals. The distribution of the plural marker influences the distribution of the NCM through the *s*-plural generalization. As such, the selection of the NCM is indirectly affected by the phonological properties of the non-head. We have further seen that conceptual properties of the non-head determine the selection of the NCM as well. Relevant factors are notions such as animacy, masshood, granularity, ... The class of a noun has an effect on the selection of the NCM and erroneous references. Finally, I have pointed out that affixes select their own NCM. The morphological make-up of the non-head is thus relevant as well when it comes to selecting an NCM.

The observations support the hypothesis that NCMs are noun class markers. Cross-linguistically, noun class markers are selected via precisely the criteria presented above, i.e. the phonological, conceptual and morphological properties of the noun. Furthermore, the mere fact that the principles of assignment are mixed is widely attested as well (Aikhenvald 2000).

A promising result of this section is the fact that the NCM is analyzed as a piece of nominal inflection which is syncretic with plural marking. It opens up the possibility to capture the formal

similarity between NCMs and plural markers, avoiding the fallacious view that NCMs are plural markers. In the next section I will present a detailed analysis of this cumulative exponence.

### 3.7 Formal analysis of the cumulative exponence of NCMs and plural markers

#### 3.7.1 Introduction

In this section I will present an overview of the inventory of Dutch noun classes. On the basis of this inventory I will present a structural analysis which captures their cumulative exponence with plural markers. Throughout the analysis I will adopt the insights from Distributed Morphology. Most importantly, I assume that syntax merges feature bundles which are realized post-syntactically along the lines of the Subset Principle, which is given in (58). (Halle and Marantz 1993, Harley and Noyer 1999).

#### (58) The Subset Principle

The phonological exponent<sup>77</sup> of a Vocabulary item is inserted into a morpheme in the terminal string if the item matches all or a subset of the syntactic features specified in the terminal morpheme. Insertion does not take place if the Vocabulary item contains features not present in the morpheme. Where several Vocabulary items meet the conditions for insertion, the item matching the greatest number of features specified in the terminal morpheme must be chosen (Halle 1997:428).

#### 3.7.2 The features of Dutch noun class markers

The table in (59) contains an overview of what an inventory of Dutch noun classes might look like. The overview serves to give an impression of the patterns, it does not have the ambition to be exhaustive. A class is defined by which plural marker and the NCM is selected by its members.

(59)

Noun Class Marker	Plural Marker	
Ø	Subclass 1.1: -s Subclass 1.2: -en	<ul style="list-style-type: none"> <li>▪ mass concepts</li> </ul>
Ø	s	<ul style="list-style-type: none"> <li>▪ most roots ending in a schwa+sonorant or in an open vowel</li> <li>▪ loanwords</li> </ul>

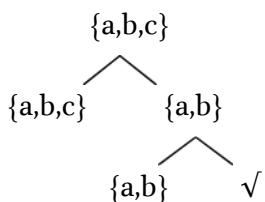
---

<sup>77</sup> A vocabulary item consists of a phonological form and information about its context of insertion. Its phonological form is called the phonological exponent of a vocabulary item.

s	s	<ul style="list-style-type: none"> <li>▪ some exceptional animate concepts (e.g. <i>kok</i> ‘cook’)</li> <li>▪ roots referring to humans ending in a schwa+sonorant or in an open vowel</li> <li>▪ diminutives</li> <li>▪ many affixes</li> </ul>
s	en	<ul style="list-style-type: none"> <li>▪ concepts with salient subparts</li> <li>▪ many affixes</li> <li>▪ abstract concepts</li> </ul>
en	en	<ul style="list-style-type: none"> <li>▪ most saliently animate and concrete concepts, i.e. prototypical atoms)</li> </ul>

To analyze these classes structurally, I adopt some main ideas from Harbour’s (2008) analysis of noun classes and number marking. He essentially proposes that number marking is a projection on top of class marking. Both projections may contain a subset of the same feature set. For example, if there is a feature set  $\{a,b,c\}$  determining the noun classes in the languages a specific class marker and a specific number marker will both contain a subset of these features. An example of such a hypothetical derivation is given in (6o).

(6o)



In the structure which is shown in (6o) the first merger of the features  $\{a,b,c\}$ , in this case the one that contains  $\{a,b\}$ , define the noun class, the second merger of this feature set derives a plural interpretation. LF thus interprets the second merger of class features as plurality. Plural marking should then not be understood as a specific feature such as [plural] which merges on top of class marking, but rather as an extra projection which influences the features values of the class projection. In short, plural marking is understood as a class extension on top of bare class marking. Bare noun class marking, which solely consists of a class projection, is thus structurally smaller than plural marking, which consists of both a class and number projection. I would like to point out that an analysis of plurality along these lines may capture the fact that plurality has both lexical and functional features. Semantically, plural marking may be subject to idiomatic interpretations (see Acquaviva 2008 and Borer 2013a). If we assume that plural marking is in fact class marking, its lexical nature comes for free. At the same time, plural marking is regular from a functional point of view in the sense that it invariably combines with count superstructure. Plurality is hard-wired in the

structure by being a functor on noun classes. In other words, the precise features of the classes are lexical, the class functor is a structural (hence, functional) property.

If one takes a closer look at the table in (59), one may observe a matryoshka doll effect in Dutch noun classes, which is summarized below. The licit combinations of NCMs and plural markers in a class are summed up in (61), the illicit ones are given in (62).

- (61) a. the NCM  $\emptyset$  can be contained in the plural marker  $-s$
  - b. the NCM  $\emptyset$  can be contained in the plural marker  $-en$
  - c. the NCM  $-s$  can be contained in the plural marker  $-s$
  - d. the NCM  $-s$  can be contained in the plural marker  $-en$
  - e. the NCM  $-en$  can be contained in the plural marker  $-en$
- (62) a. the NCM  $-s$  cannot be contained in a plural marker  $-\emptyset$ <sup>78</sup>
  - b. the NCM  $-en$  cannot be contained in a plural marker  $-\emptyset$
  - c. the NCM  $-en$  cannot be contained in the plural marker  $-s$

Through this Russian doll effect it is clear that the vocabulary item  $-\emptyset$  is smaller than  $-s$ , which is smaller than  $-en$ . Let us express this difference in size formally by assuming that the bigger vocabulary items realize more features. This idea is expressed in (63).

- (63) a.  $/\emptyset/ \leftrightarrow \{a\}$
- b.  $/s/ \leftrightarrow \{a,b\}$
- c.  $/ən/ \leftrightarrow \{a,b,c\}$

Note that the vocabulary items in (63) may realize both NCMs and plurality, even though bare noun class marking and plurality differ structurally. As such, we are able to derive the formal identity between NCMs and plural markers without the fallacious assumption that NCMs are plural markers. Having established the inventory of noun classes and the feature composition of the class markers, we are now able to derive the noun class marking and the syncretic plural marking of each class.

### 3.7.3 Deriving the classes

#### 3.7.3.1 Class 1

In this section I will discuss the precise derivation per class. Let us start with class 1. Now recall that noun classes are defined by the NCM and the plural marker they select. The first noun class can then be defined as in (64).

- (64) Class 1 is the class of roots which select  $\emptyset$  as their NCM and  $s$  as their plural marker.

It thus contains roots which select  $\emptyset$  if they are the non-head of a compound, as in (65), and which select  $-s$  as their plural marker, as in (66).

---

<sup>78</sup>This generalization follows from the fact that there is no plural marker  $-\emptyset$ .

We observe the NCM  $\emptyset$  in a compound, which is feature-wise the simplest morpheme. Let us therefore assume this class selects the simplest noun class marking possible, which is the feature set  $\{a\}$ . This derivation is shown in (67).

- (67) 

Given the inventory of vocabulary items in (63), the best candidate to realize this bare noun class marker is the one in (63)a, which is repeated in (68).

- (68)  $|\emptyset| \leftrightarrow \{a\}$

To derive the plural of this class, we merge yet another set of class features, as we have defined plural marking as a class extension above. Given that the plural of this class is expressed by the smallest vocabulary item which can express plurality, i.e. -s, I assume class is extended minimally. I therefore propose that the feature set {a,b} merges on top of the bare noun class marking. The result is shown in (69).

- (69)

```

graph TD
    A["{a,b}"] --- B["{a,b}"]
    A --- C["{a}"]
    B --- D["{a}"]
    B --- E["\sqrt{ }"]
  
```

Let us now assume that two adjacent terminal nodes of which the features stand in a subset relation to one another are unified before vocabulary insertion. This stands to reason as noun class marking and plural marking are commonly subject to cumulative exponence cross-linguistically (Kihm 2005). Postulating that plural marking is a class extension which universally triggers unification with bare class marking in case of a subset relation may capture this observation. Such unification can be derived through head movement at syntax or a PF operation of morphological merger. In any case, what will be read by vocabulary insertion is the union (in set-theoretical terms) of the two feature sets. In other words, vocabulary insertion seeks for a vocabulary item to realize the feature set in (70).

$$(70) \quad \{a\} \cup \{a,b\} = \{a,b\}$$

It will then insert the vocabulary item *-s*, which is given in (71).

- (71) /s/  $\leftrightarrow \{a,b\}$

The merger of class and number thus absorbs the feature set of the bare class marking through the union of their features sets. Cumulative exponence is derived. We have now derived the bare noun class marking and the plural marking of class 1. Below we will derive the other classes in a similar fashion.

### 3.7.3.2 Class 2

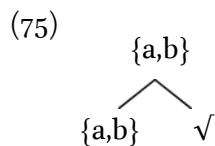
Class 2 is defined in (72). An example of one of its members as the non-head in a compound is given in (73), (74) shows this root in a plural context.

- (72) Class 2 is the class of roots which select *s* as their NCM and *s* as their plural marker.

- (73) dokter-s-briefje  
doctor-NCM-note  
'doctor's note'

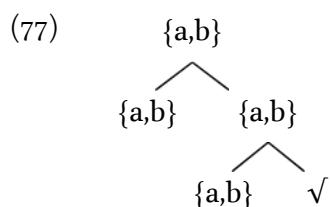
- (74) twee dokter-s  
two doctor-PL  
'two doctors'

Given that its bare noun class is realized by means of the vocabulary item *-s*, we have to assume it is marked by at least the features  $\{a,b\}$  according to the Subset Principle. The structure of the compound's non-head is given in (75). The terminal node will be realized by means of the vocabulary item in (76).



- (76) /s/  $\leftrightarrow \{a,b\}$

The plural marker is identical to the bare noun class marker. I propose the exact same feature set merges to derive plurality, as in (77). Vocabulary insertion then seeks a vocabulary item to match the feature set in (78). The best candidate is again the vocabulary item in (79).



$$(78) \quad \{a,b\} \cup \{a,b\} = \{a,b\}$$

$$(79) \quad /s/ \leftrightarrow \{a,b\}$$

We have now derived class 2.

### 3.7.3.3 Class 3

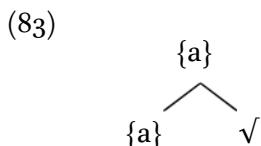
The definition of class 3 is given in (80). Its members are *en*-plurals which refer to mass concepts. An example of the bare noun class marking is given in (81), (82) shows an example of plural marking.

(80) Class 3 is the class of roots which select  $\emptyset$  as their NCM and *-en* as their plural marker.

(81) wijn- $\emptyset$ -fles  
 wine-NCM-bottle  
 'wine bottle'

(82) twee wijn-en  
 two wine-PL  
 'two kinds of wine'

I propose it selects the simplest feature set {a} as its bare class marking, on a par with the mass concepts of class 1. At the level of the bare noun class marking, mass concepts of class 1 and class 3 are thus structurally identical, which is a desirable result. The derivation is illustrated in (83).



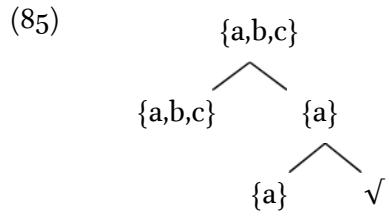
The vocabulary which will realize this node is given in (84).

$$(84) \quad / \emptyset / \leftrightarrow \{a\}$$

The plural marking of this class is realized by means of the biggest vocabulary item feature-wise, i.e. *-en*, which is given in (87). I therefore assume that the terminal node of the plural class extension is rich structurally, as in (85). Recall that the vocabulary item in (87) does not realize the number projection directly, but the unified feature set of class and number, which is given in (86).<sup>79</sup>

---

<sup>79</sup> One may wonder why syntactic plurality is derived through various features sets (e.g. [a,b][c][d][e][a,b,c]). Recall that plurality is defined above as class extension: syntactic plurality is defined as the mere merger of a second set of class features. The precise inventory of these features is class dependent. The features themselves are thus not the essence of syntactic plurality, the class extension itself is syntactic plurality.



(86)  $\{a\} \cup \{a,b,c\} = \{a,b,c\}$

(87) /ən / ↔ {a,b,c}

We have now derived the cumulative exponentence of class and number of class 3.

### 3.7.3.4 Class 4

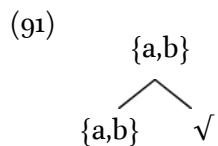
The definition of class 4 and an example of its members are given below.

(88) Class 4 is the class of roots which select -s as their NCM and -en as their plural marker.

(89) dorp-s-café  
village-NCM-pub  
'pub in a village'

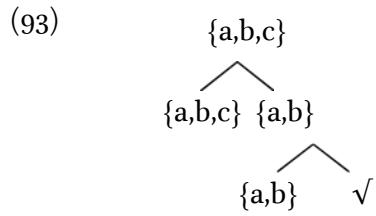
(90) twee dorp-en  
two village-PL  
'two villages'

The ingredients of its bare class marking are given in (91) and (92).



(92) /s/ ↔ {a,b}

Its cumulative exponentence with the plural marker -en is derived as in (93) and (94). The vocabulary item which realizes this cumulative exponentence is given in (95).



(94)  $\{a,b\} \cup \{a,b,c\} = \{a,b,c\}$

(95) /ən /  $\leftrightarrow \{a,b,c\}$

We have now derived class 4.

### 3.7.3.5 Class 5

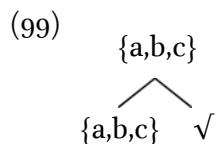
Class 5 is defined in (96). (97) and (98) contain an example.

(96) Class 4 is the class of roots which select -en as their NCM and -en as their plural marker.

(97) kat-en-staart  
cat-NCM-tail  
'cat tail'

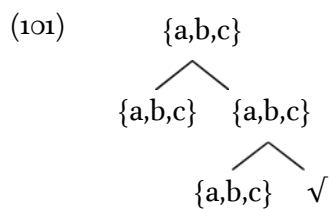
(98) twee kat-en  
two cat-PL  
'two cats'

Class 5 is the class with the richest feature set in its bare noun class marking. This is shown in (99). Resultatively, it is realized by means of the vocabulary item in (100).



(100) /ən /  $\leftrightarrow \{a,b,c\}$

Its plural marking is identical to its bare noun class marking. This is expressed below.



(102)  $\{a,b,c\} \cup \{a,b,c\} = \{a,b,c\}$

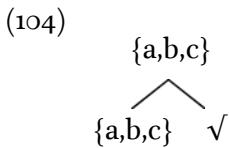
$$(103) /ən / \leftrightarrow \{a,b,c\}$$

We have now derived all attested combinations of bare noun class marking and plural marking. In the next section I will demonstrate that the unattested ones are ruled out by the analysis.

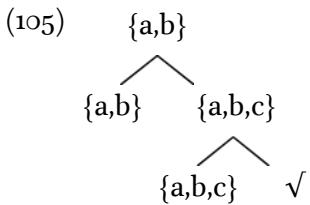
### 3.7.3.6 The s-plural generalization

We have seen that not all combinations of noun class markers are possible. More specifically, the analysis should capture the s-plural generalization (see section 5): if a root takes -s as its plural marker, it will not take -en as its NCM in the same idiolect.

Let us assume first that the feature which are associated with the NCM -en merge as the bare noun class features on top of the root. This derivation is shown in (104).



Let us now consider what happens if we try to merge an -s plural marking feature on top of it. We have seen above that the plural marker -s realizes the feature set {a,b}, as in (105).



These two terminal nodes are merged morphologically, which we have described formally as the union of the two features sets, as illustrated in (106).

$$(106) \{a,b,c\} \cup \{a,b\} = \{a,b,c\}$$

In this case the union results in the set {a,b,c}. Vocabulary insertion will now seek an appropriate candidate to realize this set. It finds the vocabulary item -en, which is repeated in (107).

$$(107) /ən / \leftrightarrow \{a,b,c\}$$

We have now derived the fact that the NCM -en can only be associated with the plural marker -en, a desirable result.

### 3.7.4 The invisibility of noun class marking in singular nominal groups

The erroneous reference facts (see section 6.4) which were discussed in section 6.4 point at the presence of noun class marking in the singular DP. After all, a head can only be involved in agreement patterns if it is indeed present in the structure. Yet, no visible noun class marking is ever present in the singular DP. How to understand this fact?

Let us go back to what happens in the plural DP. In the plural DP, noun class marking is only observable indirectly: the plural number morpheme builds on the noun class marking. We could now assume that the singular number morpheme does exactly that as well: it adds a feature on top of the noun class marking. Let us call this feature [sg].

The reader will notice that I am treating singularity not on a par with plurality. I defined plurality as a class extension, whereas I am defining singularity as a feature. Firstly, it has been noted that plurality may have quite a lexical nature in languages: plurality may show word-dependent irregularities (Acquaviva 2008). It has further been noted in the literature that singularity functions differently than plurality syntactically. According to Borer (2005:111) the plural serves to divide mass but fails to count the resulting divisions. Indeed, plurality may refer to anything that is greater (or smaller!) than one. The singular, in contrast, is both a divider and an exact counter. Given these considerations, deriving the singular and the plural somewhat differently in the syntax is a desirable result.

I will assume again that the noun class head incorporates into the number head and they are spelled out together, just as is the case in the plural DP, resulting in the following possible sets: {a,sg}, {a,b,sg}, {a,b,c,sg}. According to the Subset Principle, the following Vocabulary Items would all be candidates to spell out these sets:

- (108)     $\emptyset \leftrightarrow \{a\}$
- $\emptyset \leftrightarrow \{\text{sg}\}$
- $/s/ \leftrightarrow \{a,b\}$
- $/ən/ \leftrightarrow \{a,b,c\}$

For some sets, there would be competing subsets. For example, {a}, {sg} and {a,b} are all subsets of {a,b,sg}. Of all these subsets, {a,b} is the richest one. Yet, singularity is not spelled out a -s in Dutch, but by means of a zero affix. I am forced to assume that there is an underlying feature hierarchy that guarantees that the feature [singular] trumps the noun class features *a*, *b* and *c* (cf. Noyer 1992).

With the short proposal above I have hinted at a direction the analysis could take. I leave a more careful proposal to further research.

### 3.7.5 The values of the features *a*, *b* and *c*.

One may now wonder whether it is possible to derive what the features *a*, *b* and *c* stand for (cf. Harbour 2008 for Kiowa). I think, indeed, that it is possible to assign a core meaning to these features in the sense that prototypical countability seems to increase with the features *b* and *c*:

- (1) -en is associated with concretes and animates (prototypical atoms)
- s is associated with abstracts and inanimates (prototypical non-atoms)
- $\emptyset$  is associated with items which are prototypically not pluralized (non-augmentable concepts)

It is well-known that a high degree of animacy is linked to a high degree of individuation: if something is alive, it cannot be grinded. The feature *a* is thus probably associated with a notion that expresses a lower degree of individuation, such as grouphood, masshood or non-atomicity. The feature *b* is the sole feature which is present in all plural derivations, even in the plural marking of the default group in class 1. It therefore stands to reason this feature expresses plurality or

augmentedness (Harbour 2008). Feature *c* is associated with a high degree of individuation, concreteness and countability.

However, these properties are not consistently observable. Consider class 1, for example. This class contains various members, including mass concepts and non-human concepts. If we simply assigned a [mass] feature to the bare noun class marking of this class, we would falsely derive the fact that all its members are syntactically mass. They are thus not functional features that compositionally force the interpretational module to assign a strict meaning to the referent.

One may wonder whether a single noun can flexibly be assigned to various classes. Some data indeed seem to point in that direction. Consider the lexical items *vis* 'fish' and *kip* 'chicken', which are as common in count readings (referring to the animal) as in mass readings (referring to the meat). They do show some interesting variation. Nouns that refer to body parts of the animal or that clearly refer to a concept associated with the animal seem to get the linking element *-en*:

- (2) *vissenbek* 'fish mouth', *vissenbloed* 'fish blood', *vissenhuid* 'fish skin', *vissenleer* 'fish leather', *vissenstaart* 'fish tail', *vissennoog* 'fish eye', *vissenkom* 'fish bowl', *vissenei* 'fish egg', *vissengeheugen* 'fish memory', *vissengeslacht* 'fish gender', *vissennaam* 'fish name'
- (3) *kippenbout* 'chicken leg', *kippenborst* 'chicken breast', *kippenei* 'chicken egg', *kippenbot* 'chicken bone', *kippenbloed* 'chicken blood', *kippenhart* 'chicken heart', *kippenkarkas* 'chicken carcass', *kippenlever* 'chicken liver', *kippenkweek* 'chicken breeding', *kippenhok* 'henhouse', *kippenvoer* 'chicken feed'

Dishes that involve the meat seem to get no overt linking phoneme:

- (4) *vissoep* 'fish soup', *vispaté* 'fish pâté', *visrestaurant* 'fish restaurant', *visschotel* 'fish dish', *viswijn* 'fish wine'
- (5) *kipcurry* 'chicken curry', *kipfilet* 'chicken fillet', *kipnugget* 'chicken nugget', *kipcassoulet* 'chicken cassoulet', *kipfricassee* 'chicken fricassee', *kipgerecht* 'chicken dish', *kiphapje* 'chicken snack', *kipkroket* 'chicken croquette'

However, these tendencies are not exceptionless:

- (6) *kippenfond* 'chicken stock', *kippensoep* 'chicken soup', *kippenmousse* 'chicken mousse'

Recall that we have two methodological problems. Firstly, I have pointed out that the vocabulary inherited compounds from the past which may have preserved irregularities that may not be part of a productive process in contemporary Dutch. It is thus hard to say what the status of the exceptions in (107) is. Secondly, it is difficult to distinguish between compounds of the root primary compound type and the nominal primary compound type with a null NCM, as discussed in section 4. The data are in line with the hypothesis that *vis* 'fish' and *kip* 'chicken' select a different class when referring to meat than when referring to the animal, but admittedly, they do not conclusively show it.<sup>80</sup> Experimental linguistics could probably shed light on this issue in the future.

---

<sup>80</sup> Erroneous reference strongly suggests that *vis* 'fish' is flexible when it comes to noun class: the erroneous reference is restricted to the mass (meat) reading.

I would like to point out that nothing contradicts the assumption that the feature *a* is the feature ‘noun’ (i.e. little *n*) itself. Readers who would like to adopt this conclusion are welcome to do so. It would be in line with Borer’s (2005) proposal that the mass reading is the default reading for the nominal group.

To conclude, it is clear that the feature *a* is associated with a notion that is ranked low on a scale of individuation, whereas the feature *c* is to be associated with a notion ranked high on this scale. The feature *b* seems to be associated with plurality. However, one should keep in mind that these associations are rather loose. Whether lexical items can flexibly select various noun classes is an open question, but the data suggest it is a valid hypothesis.

### 3.8 Class and gender co-exist in Dutch

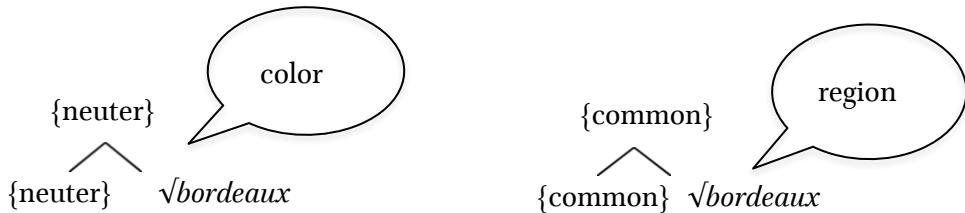
Dutch is usually not analyzed as a language with noun class marking, but it is well-known that it has gender marking. Northern Dutch has common and neuter gender and the Southern Dutch-speaking regions have the older three-way distinction (masculine, feminine and neuter). In Standard Dutch the gender of the noun is only noticeable via agreement on the adjective and the definite determiner in singular nominal groups. There is thus no doubt that singular nominal groups contain a gender head, as it has noticeable syntactic effects. Plural nominal groups will get default common gender. Yet, one may assume the original lexical gender is also contained in the plural nominal group for the following reason.

As in many grammatical gender systems, Dutch gender is assigned to animates mostly in ways that make sense biologically, but stored as an arbitrary feature in the lexicon for most other nouns. However, certain semantic classes do select gender consistently. For example, all metals, colors and cities are neuter. All morphologically simplex names for vegetables and fish are common and so are names for French regions, rivers, etc.

- (7) het ijzer/goud/zilver/kwik/lood/koper/tin/...  
 ‘the<sub>NEUTER</sub> iron/gold/silver/mercury/lead/copper/tin/...’
- (8) het rood/blauw/geel/groen/oranje/paars/wit/zwart/magenta/cyaan/scharlaken/ bordeaux  
 ‘the<sub>NEUTER</sub> red/blue/yellow/green/orange/purple/white/black/magenta/cyan/  
 scarlet/burgundy’
- (9) de tomaat/spinazie/sla/kool/erwt/wortel/paprika/...  
 ‘the<sub>COMMON</sub> tomato/spinach/lettuce/cabbage/pea/carrot/bell pepper/...’
- (10) de aal/rog/baars/paling/haring/ansjovis/kabeljauw/zalm/barbeel/  
 forel/haai/wijting/karper/leng/brasem/braam/sardine/...  
 ‘the<sub>COMMON</sub> eel/ray/perch/eel/herring/anchovy/cod/salmon/barbel/  
 trout/shark/whiting/carp/ling/bream/sea-bream/sardine/...’
- (11) de Loire/Beaujolais/Champagne/Bordeaux/Bourgogne/....  
 ‘the<sub>COMMON</sub> Loire/Beaujolais/Champagne/Bordeaux/Bourgogne/....’

Gender may facilitate the assignment of meaning in the Encyclopedia. In this way, gender is a classifying system in its own right. Interestingly, gender differences may correlate with meaning differences:

- (12) a. het bordeaux  
           the<sub>NEUTER</sub> burgundy  
           (the color)
- b. de Bordeaux  
           the<sub>COMMON</sub> Bordeaux  
           (the region)



The merger of various lexical genders may thus correspond to meaning differences. The meaning differences can be attested in plural nominal groups as well:

- (13) Frankrijk heeft geen twee Bordeaux, hoor!  
     France has no two Bordeaux, PRT  
     ‘France really does not have two Bordeaux! There is only one region with that name.’

- (14) De stylist combineerde twee verschillende bordeaux.  
     the stylist combined two different bordeaux  
     ‘The stylist combined two different shades of burgundy.’

Given that lexical gender is responsible for the meaning differences, it must be the case that the plural nominal groups projects lexical gender as well, even though it does not surface in the overt realization. Proposals in Distributed Morphology on other languages have proposed that gender may be subject to Impoverishment (Bonet 1991, Noyer 1992) through markedness in plurals: gender features are banned in the marked context of plurality (Bobaljik 2002; Bailyn and Nevins 2008, Harley 2008; Nevins 2011, Kramer 2019).<sup>81</sup>

Finally, we may wonder whether gender is present in the compound's non-head next to gender marking. The West-Flemish compounds, which show feminine gender expressed overtly by means of an affix, suggest that this is indeed the case. We can now formulate an overview of which projections are projected in which context:

---

<sup>81</sup> Note that the mere fact that gender may be realized in the nominal compound (see the West-Flemish data in section 4) is then a further indication that the nominal compound does not contain a plural projection.

(15)

	singular DP	plural DP	compound's non-head
bare class	✓	✓	✓
extended class	X	✓	X
lexical gender	✓	✓	✓
determiner-layer	✓	✓	X

The projections that are present in the compound's non-head are those projections that serve to identify the concept the lexical item refers to: noun class marking and gender marking. The projections that are absent in the non-heads compound are those projections that connect to a referent: number marking and a determiner layer<sup>82</sup>.

### 3.9 Conclusion

I started this contribution with pointing out that so-called linking elements may actually fulfil two very different roles: they may be phonotactic elements or they may realize a morphosyntactic head. In case they are merely phonotactic elements, the left-hand part of the compound may be a bare root, as pointed out by Nóbrega (2020), which is an interesting domain of study, as it is a rare context in which we observe a bare root (De Belder 2017). If they are morphosyntactic elements, the data are no less interesting: they tell us more about the functional heads that play a role in the language.

I presented a case study of linking elements in Dutch which do fulfil a morphosyntactic role. I started with pointing out that when 'linking elements' are of this type, the linguist should be aware of the fact that they may be null morphemes. I argued that, indeed, Dutch may have such a null affix in its inventory. Further research questions that are relevant for this type of linking elements are which category the linking element belongs to and which heads it realizes.

I argued that the Dutch linking elements are in fact noun class markers and as such, I derived the fact that two of them are identical to plural marking through cumulative exponence. The analysis shows that a careful study of such elements provides us with new insights into two domains simultaneously. Firstly, we learn more about the morphosyntax of the compound's left-hand part, the structure of the compound itself and the inventory of compounding types in the language. The importance of the left-hand part cannot be underestimated in my opinion. It is a unique domain: it shows how a category behaves functionally in the absence of functional material that guarantees reference, such as number and the D-layer. As such, we learn, secondly, something about the morphosyntax of a category. The Dutch nominal domain has nominal classes which are less visible in the nominal group as they are syncretic with number marking. Take away number marking and they suddenly become visible.

---

<sup>82</sup> I did not discuss the determiner layer in this contribution, but it is a well-known property of a compound's non-head that it does not contain a D-layer (see De Belder and Van Koppen 2016 and references therein).

It would be interesting to learn whether, cross-linguistically, linking elements that fulfil a morphosyntactic role are restricted to the lower functional domain of the noun. Are there languages that allow adjectival, verbal or prepositional layers? Within the nominal domain, which projections exactly are identified in this context? I hope this case study has shown that linking elements fully deserve the attention of the theoretical linguist in collaboration with the experimental linguist.

## 4 Movement to D/d in Compounds and Phrases

### Referential compounding, -s possessives and title expressions in Dutch

#### Abstract

It is generally assumed that non-heads of compounds cannot be referential: they are assumed to lack a D-layer. In this article I argue that this assumption is false. Dutch has compounds with a referential non-head. These non-heads belong to the types of lexemes Longobardi (1994) identified as typically subject to N-to-D raising: proper names, kinship names, names of days of weeks and Bank holidays, unique entities and contextually unique entities. I argue that the compounds contain a nominalizing head specified for uniqueness and referentiality, which I call little d. The non-head of the compound incorporates into little d. I compare the morphosyntax of these non-heads to the morphosyntax of modificational compounds, of proper names in argument position, of the Dutch -s possessive and title expressions, which I argue are Dutch referential construct states.

#### 4.1 Introduction

This article presents a newly discovered type of compounding in Dutch, which, I argue are compounds with referential non-heads as exemplified in (1):

- (16) zon-s-hoogte  
sun-s-height  
'solar altitude'

I will contrast this type of compounding with, what I will argue are referential construct states, as in (17), and -s possessives, as in (18):

- (17) kanselier Merkel  
chancellorMerkel  
'chancellor Merkel'
- (18) Marie-s hypothese  
Mary-s hypothesis  
'Mary's hypothesis'

The referential compound is the fourth logical combination in Borer's (2011) classification, but it was assumed to be non-existent. Borer (2011) argues, on the basis of Hebrew data, that N-N concatenations could either be constructs or compounds: constructs do not show syntactic head incorporation, whereas compounds do. Furthermore, the non-head could be either modificational or referential, a semantic distinction which is derived from a syntactic difference: modificational non-heads occupy a classifier position in my inventory of Dutch compounds (see chapter 3), the referential ones include a determiner layer. Of the four possible combinations three were argued to

be realised in Hebrew: there is the modificational compound, the modificational construct and the referential construct. I aim to show that the fourth logical possibility occurs in Dutch. Compounds of the type *zonshoogte* ‘solar altitude’ are referential compounds and thus realise the fourth possible cell in table 1:

Table 1: an inventory of possible N-N concatenations

	modificational non-head	referential non-head
<b>construct</b>	m-construct (Borer 2011)	r-construct (Borer 2011) & Dutch title expressions (see below)
<b>compound</b>	(m)-compounding (Borer 2011) & Dutch nominal compounding (see chapter 3)	r-compounding: <i>zonshoogte</i> (this chapter)

The article is structured as follows. Section 2 contrasts the referential compound with other types of primary compounding in Dutch presented in chapters 2 and 3 of this bundle. Section 3 argues that the non-head of the referential compound indeed has referential properties. Section 4 contrasts the compound with the -s- possessive and section 5 contrasts it with a referential construct state in Dutch.

#### 4.2 Referential compounds versus other Dutch primary compounding types

De Belder (2017) and (to appear) (i.e. chapters 2 and 3 in this bundle) identify two major Dutch primary compounding types: those of which the non-head is a bare root (De Belder 2017) and those for which the non-head merges with a classifier head (De Belder to appear).<sup>83</sup> The non-head of the bare root compound can be associated with any category and never merges with an overt ‘linking element’:

- |                    |                 |
|--------------------|-----------------|
| (19) a. speur-hond | b. drie-luik    |
| track-dog          | three-panel     |
| ‘tracking dog’     | ‘triptych’      |
| c. snel-trein      | d. achter-grond |
| fast-train         | back-ground     |
| ‘high-speed train’ | ‘background’    |

The non-head of the classifier type merges with a nominal classifying projection. It thus invariably gets a nominal interpretation. This type, I believe, is the Dutch version of what Borer (2011) identified as genuine compounding in Hebrew. The type is easily recognised by the presence of a so-called linking element -s- or -en-:

---

<sup>83</sup> I also described ANN-compounds (an instance of non-primary compounding) together with Marjo van Koppen (De Belder & Van Koppen 2016), see chapter 5.

- (20) a. peer-en-boom  
       pear-NCM-tree  
       'pear tree'
- b. varken-s-hok  
       pig-NCM-pen  
       'pig's pen'

I argued that this 'linking element' (a misnomer!) realises the functional head of nominal classification. Furthermore, I pointed out that nominal classificational heads can in principle be realised by null morphemes as well and I argued there is actually dialectal evidence that this indeed happens in Dutch, probably also in the standard language. The reasoning is rather complex and lengthy, so I refrain from repeating it here (see chapter 3). According to the reasoning described there, the following compounds could be relevant examples (De Belder to appear).

- (21) a. siroop-fles  
       syrup-bottle  
       'syrup bottle'
- b. klei-grond  
       clay-soil  
       'clay soil'
- c. wol-draad  
       wool-yarn  
       'wool-yarn'

As a result, the classifier head, I argued, can be realised by means of  $\emptyset$ , -s- [s] or -en- [ə(n)]. Compounds of this type with a null marker are at the surface of course indistinguishable from compounds with a bare root as their non-head. It is an unlucky empirical consequence.

The non-head of the classifying (i.e. modificalional) compounding type selects the precise realisation of the classifier head. One can thus expect a quite regular selection between the non-head's lexeme and its classifying exponent of choice ( $\emptyset$ , -s- or -en-) (see chapter 3 for a more nuanced discussion):

- (22) a. kat-en-luik            b. kat-en-voer            c. kat-en-staart            d. kat-en-bak  
       cat-NCM-shutter        cat-NCM-food        cat-NCM-tail        cat-NCM-box  
       'cat flap'                'cat food'            'cat tail'            'cat litter box'
- (23) a. ezel-s-dracht            b. ezel-s-bruggetje            c. ezel-s-oor  
       donkey-NCM-pregnancy        donkey-NCM-bridge.DIM        donkey-NCM-ear  
       'long pregnancy'                'mnemonic'                'dog-ear'

The present article, however, draws attention to the fact that sometimes non-heads occur with an -s-, even though they would typically be restricted to bare root compounding or select zero marking or -en- as a classifying compound. I argue that when these non-heads are bare roots or when they select their typical exponent of the classifying head (here zero or -en-) they are modificalional compounds, if they select the unexpected -s-, they are referential compounds. This results in the following inventory for Dutch compounds:

Table 2: an inventory of Dutch primary compounding

	bare root compounding	noun class marking compounding	referential compounding
‘linking element’?	none: the non-head is a bare root	$\emptyset$ , [s] or [ $\theta(n)$ ]	[s]
modificational or referential?	modificational		referential

These three types are illustrated in the following examples:

- (24) a. *kreeft-woord*  
          lobster-word  
          ‘palindrome’      b. *kreeft-en-soep*  
          lobster-EN-soup  
          ‘lobster soup’     c. *Kreeft-s-keerkring*  
          lobster-s-tropic  
          ‘Tropic of Cancer’

Example (24)a shows a bare root compound: the root *kreeft* ‘lobster’ is not followed by any material, i.e. the root is bare. Example (24)b shows a noun class marking compound: the root *kreeft* ‘lobster’ is followed by its usual noun class marker which surfaces as *-en*. Example (24)c shows a referential compound, in which the root *kreeft* ‘lobster’ is followed by *-s-*, which characterises this type of compounding.

Note that referential compounds can be recognised because an *-s-* appears where it is not immediately expected. I now discuss this in some more detail. De Belder (to appear, chapter 3) argued that non-heads with the semantics of kinship names and proper names are probably always bare root compounds:

- (25) a. *moeder-melk*      b. *moeder-taal*  
          mother-milk        mother-language  
          ‘breast milk’      ‘mother tongue’
- c. *vader-beeld*  
     father-image  
     ‘conception of the father figure’
- d. *vaderbeweging*  
     father-movement  
     ‘movement for the emancipation of divorced fathers’
- f. *Pieter-baas*      e. *Pieter-man*  
     Peter-boss        Peter- man  
     ‘black Pete’      ancient coin with the image of Saint-Peter/  
                           name of a certain fish (*Trachinus draco*)

However, one does find instances of kinship names and proper names followed by an *-s-* in Dutch compounding:

- |                                     |                               |                  |
|-------------------------------------|-------------------------------|------------------|
| (26) a. moeder-s-kind               | b. vader-s-zijde              | c. Pieter-s-zoon |
| mother-s-child                      | father-s-side                 | Peter-s-son      |
| 'child too dependent on the mother' | 'father's side of the family' | (family name)    |

I will argue that these compounds are referential compounds.

Similarly, the Dutch roots *dag* 'day' and *jaar* 'year' typically do not select a 'linking element', either because they invariably occur in bare root compounding or because they are instances of noun class marking compounds which select a zero exponent of noun class marking (I am principally unable to tell):

- |                   |               |                        |                 |
|-------------------|---------------|------------------------|-----------------|
| (27) a. dag-deel  | b. jaar-beurs | c. jaar-balans         | d. jaar-getijde |
| day-part          | year-fair     | year-balance.sheet     | year-tide       |
| 'part of the day' | 'trade fair'  | 'annual balance sheet' | 'season'        |

Yet, again, one does find instances of exactly these roots selecting an *-s*:

- |                            |   |
|----------------------------|---|
| (28) a. verjaardag-s-taart | b. zondag-s-kind                                |
| birthday-s-pie             | sunday-s-child                                  |
| 'birthday pie'             | 'child born on a Sunday and born for good luck' |
| c. nieuwjaar-s-feest       |   |
| new.year-s-party           |   |
| 'New Year's party'         |   |

Again, the claim is that they are referential compounds.

Then there are noun class marking compounds of which the non-head selects *-en-* to realise the noun class marking:

- |                                    |                   |                  |                    |
|------------------------------------|-------------------|------------------|--------------------|
| (29) a. zonn-en-bank <sup>84</sup> | b. kreeft-en-soep | c. naam-en-lijst | d. maan-en-stelsel |
| sun-EN-bench                       | lobster-EN-soup   | name-EN-list     | moon-EN-system     |
| 'tanning bed'                      | 'lobster soup'    | 'list of names'  | 'moon system'      |

Yet, again, one does find instances of these roots selecting an *-s*, which I claim to be instances of referential compounding:

- |                      |                       |                     |                         |
|----------------------|-----------------------|---------------------|-------------------------|
| (30) a. zon-s-hoogte | b. Kreeft-s-keerkring | c. naam-s-wijziging | d. maan-s-verduistering |
| sun-s-height         | Cancer-s-tropic       | name-s-change       | moon-s-eclipse          |
| 'solar altitude'     | 'Cancer Tropic'       | 'name change'       | 'lunar eclipse'         |

To be entirely clear, I do not aim to argue that *all* instances of Dutch compounds with an *-s* are referential compounds. As can be deduced from table 2, I maintain the claim that the *-s* can be a

---

<sup>84</sup> The official spelling would be *zonnenbank*.

realisation of a nominal classifying head as well.<sup>85</sup> As a consequence, I believe it is probably so that there are compounds which select both the -s- as modificational compounds and as referential compounds. It is another instance of an unlucky homophony in Dutch compounding.

#### 4.3 The referentiality of the non-head

The non-referentiality of the non-head is often taken to be a defining criteria of compounding (see e.g. Borer 2005a:84; Borer 2011), so claiming that it can be referential is quite controversial. Then, why would I think that, indeed, I have discovered instances of compounds with a referential non-head? In this section I present arguments in favour of the referentiality of the non-head, in the next section I present arguments in favour of the compounding status.

Consider the following four arguments for the referentiality of the non-head. Firstly, the non-heads that occur in the referential compounds seem to belong to a specific group; they are highly reminiscent of the type of lexemes Longobardi (1994) identified as typically subject to N-to-D raising: proper names (*Pieter* 'Pete'), kinship names (*moeder* 'mother', *vader* 'vader'), names of days of the weeks (*zondag* 'Sunday') and holidays (*Nieuwjaar* 'New Year', *verjaardag* 'birthday'), and unique entities (*zon* 'sun', *Kreeft* 'Cancer (the constellation)', *maan* 'moon'). The first ingredient to become referential is thus in place: these lexemes are excellent candidates to raise to the D-layer syntactically where they can gain referential semantics.

Secondly, the referential compounding seems to imply universal uniqueness or at least uniqueness in the context. For example, *Kreeftskeerkring* 'Tropic of Cancer' uniquely refers to the Cancer constellation. This contrasts with *kreeftensoep* 'lobster soup' which does not show the -s- and which does not imply unique reference to a lobster at all. In fact, it does not even imply the presence of lobster: it is pretty easy to find a recipe for *kreeftensoep zonder kreeft* 'lobster soup without lobster' on Google. Similarly, *manenstelsel* 'moon system' does not refer to the Earth's unique moon, whereas *maansverduistering* 'lunar eclipse' does refer to the unique moon as we know it. For words such as *naam* 'name', *vader* 'father', *zondag* 'Sunday' and *verjaardag* 'birthday' the uniqueness is contextual rather than universal. This does not constitute a problem: it is known from research on definiteness that contextual uniqueness suffices for referentiality (Lyons 1999).

Thirdly, in the absence of uniqueness, the referential compound is excluded, as illustrated by the following minimal pair:

- (31) zon-s-hoogte

sun-s-height

'solar altitude'

- (32) \* ster-s-hoogte

star-s-height

Fourthly, not only the uniqueness, but the referentiality itself is implied. Compare the following contrast. The following fully acceptable dialogue illustrates the classical, familiar non-referentiality of the non-head of a modificational compound (*hondenmand* 'dog bed'):

---

<sup>85</sup> I also do not exclude the possibility that there are yet other types of compounding in Dutch which have not been discovered yet.

- (33) A: We kopen een hondenmand.  
 B: Oh, heb jij een hond?  
 A: Nee, eigenlijk niet, we gaan de mand gebruiken voor onze kat.  
 B: Ja, je hebt gelijk, dan ligt ze wat ruimer.

'A: We are buying a dog bed.  
 B: Oh, do you have a dog?  
 A: No, actually not, we are going to use the bed for our cat.  
 B: Yes, I see, it will be a bit more spacious for her then.'

For the referential compound *verjaardagsfeest*, below, however, a dialogue parallel to the one in (33) is excluded -or at least very odd- due to the fact that the existence of the birthday is actually implied:

- (34) #A: We organiseren een verjaardagsfeest.  
 B: Oh, is er een verjaardag?  
 A: Nee, eigenlijk niet, we organiseren het feest voor een huwelijk.  
 B: Ja, je hebt gelijk, dat is vast goedkoper.

'A: We are organising a birthday party.  
 B: Oh, is there a birthday?  
 A: No, actually not, we are organising the party for a wedding.  
 B: Yes, I see, it's probably cheaper.'

I conclude that the non-head of these compounds has unique reference. The consequence is that the long-held belief that the non-head of a compound is by definition non-referential may be falsified.

#### 4.4 Referential compounds versus the -s- possessive construction

The present section argues that the referential compound has true compounding status (and, as such, word-hood status) by comparing it to a phrase with similar properties: the Dutch -s possessive construction (see also De Belder 2009).<sup>86</sup>

The Dutch -s possessive resembles the English saxon genitive, but it is more restricted. In Dutch, the -s possessive construction is restricted to kinship nouns and proper names. Hence, the examples in (35) are excluded, whereas the ones in (36) are fine.

- |  |   |
|--|---|
| (35) a. * gisteren-s arrangementen<br>yesterday-s arrangements | b. * vrouw-s hypothese<br>woman-s hypothesis        |
| c. * een vrouw-s hypothese<br>a woman-s hypothesis             | d. * de vrouw-s hypothese<br>the woman-s hypothesis |

---

<sup>86</sup> I am using the term 'construction' in a descriptive, non-theoretical sense.

- (36) a. Marie-s hypothese  
           Marie-S hypothese  
           Mary's hypothesis
- b. papa-s auto  
           daddy-S car  
           'daddy's car'

When one studies the data in somewhat more detail, the generalisation is that the Dutch -s possessive construction is restricted to those nouns which can undergo N-to-D raising in argument position. Kinship nouns and proper names are of course roots that have the appropriate semantics to undergo N-to-D-raising, hence they can occur without an overt determiner in argument position, suggesting that the root itself has raised to D (Longobardi 1994):

- (37) Ik ontmoette Marie in Parijs.  
       I met           Mary in Paris.  
       'I met Mary in Paris.'
- b. Ik zag papa.  
       I saw        daddy  
       'I saw daddy'.

If the noun cannot undergo N-to-D-raising in argument position, the -s possessive construction is excluded. The contrast between the examples (38) and (39) shows that the noun (i.e. the root in the structurally nominal position) *zon* 'sun' requires an overt determiner in Dutch in argument position. Example (40) shows that the noun cannot occur as the possessor in the -s possessive construction. The proper names of rivers illustrate the same fact. They cannot occur without an overt determiner, as names of rivers need a definite article in Dutch, and they do not occur as the possessor in the -s possessive either (see section 6 for an account):

- (38) Ik zie de zon.  
       I see    the sun  
       'I see the sun.'

- (39) # Ik zie zon.<sup>87</sup>  
       I see    sun

- (40) \* zon-s zachte warmte  
       sun-S gentle warmth

- (41) Ik zie de Seine.  
       I see the Seine  
       'I see the Seine'

- (42)\* Ik zie Seine.  
       I see Seine

- (43)\* Seine-s flikkerende spiegeling  
       Seine-S flickering reflection

---

<sup>87</sup> A mass reading is possible, as in 'I see some sunlight.'

The *-s* possessive construction and the referential compound show resemblances. Their non-heads are referential and they are both marked by an *-s*. Yet, they also show significant empirical differences, showing that we are dealing with two distinct structures.

Consider the following four criteria that distinguish between the two structures. Firstly, recall the restriction that only non-heads that can undergo N-to-D movement when occurring in argument position can occur as the possessor in *-s* possessive constructions. A parallel restriction does not hold for referential compounds. The referential compound allows all roots as a non-head that have a unique, universal or contextual, reference. The illicit *-s* possessive construction in (40) thus contrasts with the licit referential compound in (44):

- (44) zon-s-verduistering  
sun-s-eclipse  
'solar eclipse'

I postpone an account for this contrast till section 6. For now it is important that the contrast exists as a criterion to distinguish between the two structures.

Secondly, the compounds qualify for word-hood in the sense that they can be lexicalised: they are stored in the native's speaker memory and in Dutch dictionaries. In that sense, a native speaker can distinguish between stored compounds, perhaps even with an idiomatic meaning, and newly formed compounds, i.e. neologisms, which are to be interpreted literally:

- (45) Ze is een zondag-s-kind.  
she is a Sunday-s-child  
'She is born on a Sunday and thus for good luck.'

- (46)# Ze is een maandagskind.  
she is a Monday-s-child  
(The speaker expresses that there is a salient connection in the discourse between the child and Mondays.)

- (47) Ze is een zondag-s-oma.  
she is a Sunday-s-grandmother  
(The speaker expresses that there is a salient connection in the discourse between the grandmother and Sundays, for example, because this grandmother only visits the family on Sundays.)

I do not fully exclude a humorous use of the neologism in (31), contrasting it with the compound in (30). However, the mere fact that it would be considered humorous illustrates the point that it is not lexicalised. The *-s* possessive construction does not qualify for word-hood in that sense: it is simply a freely generated syntactic constituent. Syntactic constituents are never experienced as 'neologisms':

- (48) moeders auto/fiets/jurk/laptop/...  
mother's car/bicycle/dress/laptop/...

Thirdly, the compounds qualify for word-hood phonologically: they receive compound stress, i.e. main stress falls on the non-head. The *-s* possessive construction, in contrast, receives the stress of a syntactic constituent:

- (49) a. 'moeder-s-kind  
mother-s-child  
'child too dependent on the mother'
- b. 'Kreeft-s-keerkring  
Cancer-s-tropic  
'Cancer Tropic'

- (50) a. moeder-s 'auto  
mother-s car  
'mother's car'
- b. Marie-s 'vriendenkring  
Marie-s friend.circle  
'Mary's circle of friends'

Fourthly, the compound qualifies for word-hood in the sense that it cannot be interrupted by other words: an intervening adjective is excluded. The *-s* possessive construction, in contrast, allows for intervening adjectives:

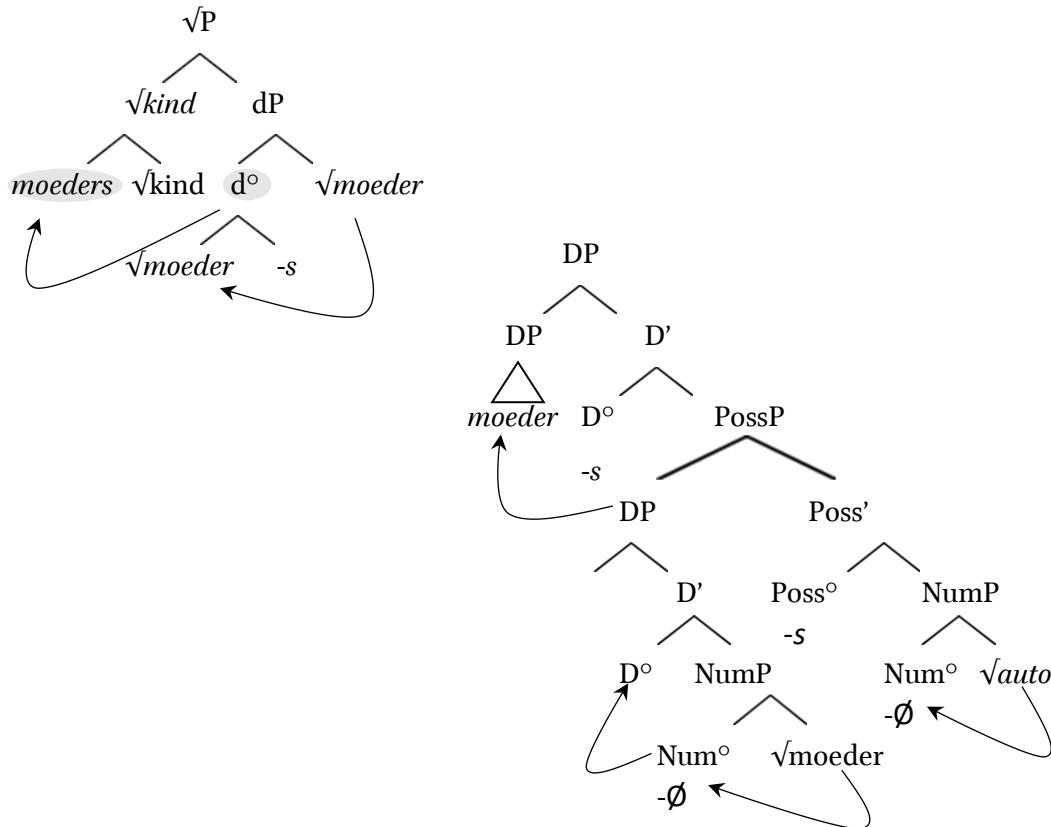
- (51) moeders   mooie   autootje  
mother's   pretty   car.DIMINUTIVE  
'mother's pretty little car'

- (52) Ze   was   mijn   grootmoeder   van   vader-s-zijde.  
she   was   my   grandmother   of   father-s-side  
'She was the grandmother of my father's side of the family.'

- (53) \* Ze   was   mijn   grootmoeder   van   vader-s   arme   zijde.  
she   was   my   grandmother   of   father-s   poor   side

I conclude that the referential compound differs from the *-s* possessive construction. The referential compound qualifies for word-hood and is thus truly a compound. Syntactically, this implies that it is derived through head incorporation, which counts as the syntactic movement that defines word-hood for compounds (Mithun 1984, Harley 2009). I propose that the non-head incorporates into a functional head called little  $D^\circ$ , which is characterised by nominality and uniqueness: [n, unique]. It surfaces as *-s-*. The non-head plus *-s-* subsequently incorporates into the head of the compound. The *-s* possessive construction, in contrast would not undergo head incorporation. The non-head rather moves to Spec,DP and the *-s* occupies the  $D^\circ$  position from its base position in what could either be analysed as Spec,nP or Spec,PossP (see Abney 1987 on the Spec,DP position for the possessor and see Radford 2000, Alexiadou, Haegeman and Stavrou 2007:568 for the base position in Spec,nP for possessors.) Corver (1990) proposes the *-s* ends in the  $D^\circ$  head.

- (54) a. moeder-s-kind  
           mother-s-child  
           'child too dependent on the mother'
- b. moeder-s auto  
           mother-scar  
           'mother's car'



#### 4.5 Referential compounds versus the referential construct state

In this section I argue that Dutch title expressions are referential construct states (see also De Belder 2009). It will become clear that they are quite different from the referential compounds.

Title expressions are illustrated in (55)-(56). They always consist of one or more bare common nouns followed by a proper name.

- (55) graaf Dracula  
       count Dracula  
       'count Dracula'
- (56) professor doctor Drosten  
       professor doctor Drosten  
       'professor doctor Drosten'

The order cannot be changed, in the sense that the bare common noun needs to precede the proper name:

- (57) \* Dracula graaf  
Dracula count

The construction needs to contain at least one proper name:

- (58) \* koningin paleis  
queen palace

The whole functions semantically as a proper name as can be deduced from the following two tests. Firstly, Dutch definite NPs allow for a generic reading, as shown in (59), except when they function as proper names, as in (60). As can be seen in (61), title expressions pattern with proper names in this respect.

- (59) De hond is een trouw dier.  
the dog is a loyal animal  
✓ generic reading: 'Every dog is a loyal animal.'
- (60) Albert is een flamboyante man.  
Albert is a flamboyant man  
\* generic reading: 'Every Albert is a flamboyant man.'
- (61) Prins Albert is een flamboyante man.  
Prince Albert is a flamboyant man  
\* generic reading: 'Every prince Albert is a flamboyant man.'

Secondly, Dutch definite NPs allow for a *de dicto* and a *de re* reading, as in (62), proper names only allow for a *de re* reading, as in (63). Again, the title expression patterns with the proper names as it only allows for a *de re* reading, as illustrated in (64).

- (62) Marie wil met haar buurman trouwen.  
Marie wants with her neighbour marry  
✓ 'Marie has the wish to marry the person living next to her, whoever that may be.' (de dicto)  
✓ 'Marie wants to marry a specific man and this man is her neighbour.' (de re)
- (63) Marie wil met William trouwen.  
Marie wants with William marry  
\* 'Marie has the wish to marry no matter which man, as long as he is called William.' (de dicto)  
✓ 'Marie wants to marry a specific man and this man is called William.' (de re)

- (64) Marie wil met prins William trouwen.  
 Marie wants with prince William marry  
 \* 'Marie has the wish to marry no matter which man, as long as he is called  
 prince William.' (de dicto)  
 ✓ 'Marie wants to marry a specific man and this man is prince William.' (de re)

I conclude that title expression are DPs which consist of one or more bare common nouns followed by a proper name and the whole functions semantically as a proper name.

One could think that the bare common noun needs to belong to a specific semantics field as it commonly refers to nobility, clergy, military ranks or professions, as shown in the following examples:

(65) prins Charles

prince Charles

'prince Charles'

(66) priester Damiaan

priest Damian

'father Damian'

(67) kapitein Von Trapp

captain Von Trapp

'captain Von Trapp'

(68) professor Curie

professor Curie

'professor Curie'

This could suggest that nouns referring to nobility, clergy, military ranks or professions form a closed class in the lexicon, which, for example, share a feature [+ title] or [+unique reference] or lexical selection properties. However, this cannot be the case, given that the construction can be applied productively. Indeed, any common noun is licit as the bare common noun in the title expression:

(69) boekenkast Billy

book case Billy

'book case Billy'

Even nonce formations can occur as the bare common noun in the title expression:

- (70) naakstrandgemeente Bredene  
nude.beach.town Bredene  
'Bredene, the town with a nude beach'

Note that the 'proper name' does not need to be stored as a proper name either, the title expression is interpretable, regardless which roots realise the syntactic positions (cf. Borer 2005a):

- (71) bedsofa Vimle  
bed.couch Vimle  
'bed couch Vimle'

This shows that the title expression does not depend on a specific property of the lexical items involved. Structural properties license the title expressions. This raises the questions which syntactic structure is realised by the title expression and why the whole functions as a proper name. I propose that the bare common noun is licensed by N°-to-D° movement in a construct state. The title interpretation is a result of this syntactic structure.

In Hebrew, a construct state is a DP which consists of a bare, unstressed head noun which is immediately followed by a genitival phrase that is not overtly case marked. A variety of semantic relations can hold between them (Cf. Borer 1984, Ritter 1991, Siloni 1997, all Hebrew examples below are taken from Siloni 1997:21-26 unless otherwise noted).

- (72) beyt ha-'is  
house the-man  
'the man's house'

Title expressions and the Hebrew construct state share structural similarities. Firstly, in both constructions prepositions cannot intervene between the head and the complement:

- (73) beyt (\*sel) ha-'is [Hebrew]  
house of the-man

- (74) professor (\*van) Einstein [Dutch]  
professor of Einstein

Secondly, in both constructions an initial determiner is illicit. This is immediately clear for Hebrew:

- (75) (\*ha)-beyt ha-'is [Hebrew]  
the-house the-man

For the Dutch title expressions, we first need a context in which a construction with a proper name would tolerate a determiner in the first place. Such examples exist: definite determiners may merge with proper names referring to males in informal, substandard Belgian Dutch.

- (76) Ik heb den Borrekens gezien. [Belgian Dutch]

I have the.MASC.SG Borrekens seen

'I have seen the man we know as Borrekens.' (*Borrekens* is a family name.)

- (77) Ik heb den Bert gezien. [Belgian Dutch]

I have the.MASC.SG Bert seen

'I have seen Bert.'

Yet, in the title expression such an initial determiner is excluded, on a par with the Hebrew construct state in (75).

- (78)\* Ik heb den baron Borrekens gezien. [Belgian Dutch]

I have the.MASC.SG baron Borrekens seen

Without the title expression, the example would have been licit, see (76) above and see(79):

- (79) Ik heb den baron gezien. [Belgian Dutch]

I have the.MASC.SG baron seen

Thirdly, in both constructions the whole DP inherits the referential properties of the second part. This is shown in (80) for Hebrew: the (in)definiteness of the second part of the construct state determines the (in)definiteness of the whole construct state (examples taken from Borer 1984:45):

- (80) ben ha-melex [Hebrew]

son DEFINITE-king

'the prince'

- (81) ben melex [Hebrew]

son king

'a prince'

For Dutch, we have seen that the entire title expression inherits the referential properties of the second part: the whole expression functions semantically as a proper name, as they resist a generic reading and a *de dicto* reading.

Fourthly, both structures are recursive:

- (82) gag beyt ha-'is [Hebrew]  
 roof house the-man  
 'the roof of the house of the man'
- (83) ingenieur doctor doctor professor erector [Dutch]  
*engineer doctor doctor professor honorary.president*  
 associatievoorzitter baron Oosterlinck  
*association.president baron Oosterlinck*  
 'engineer doctor doctor professor honorary president president of the association baron Oosterlinck'

Fifthly, for both constructions the presence of the complement is obligatory, as can be seen in (84)-(87):

- (84) beyt ha-'is [Hebrew]  
 house the-man  
 'the man's house'

- (85)\* beyt [Hebrew]  
 house

- (86) Ik heb paus Johannes Paulus II in het real gezien. [Dutch]  
 I have pope Johannes Paulus II in the real seen  
 'I have seen pope Johannes Paulus II in real life.'

- (87)\* Ik heb paus gezien. [Dutch]  
 I have pope seen

Sixthly, in both constructions, the head noun is de-stressed.

- (88)\* BAYIT ha-'is [Hebrew]  
 house the-man  
 (*bayit* is the stressed form of the unstressed *beyt*)

- (89)\* koning'in Mathilde [Dutch]  
 queen Mathilde

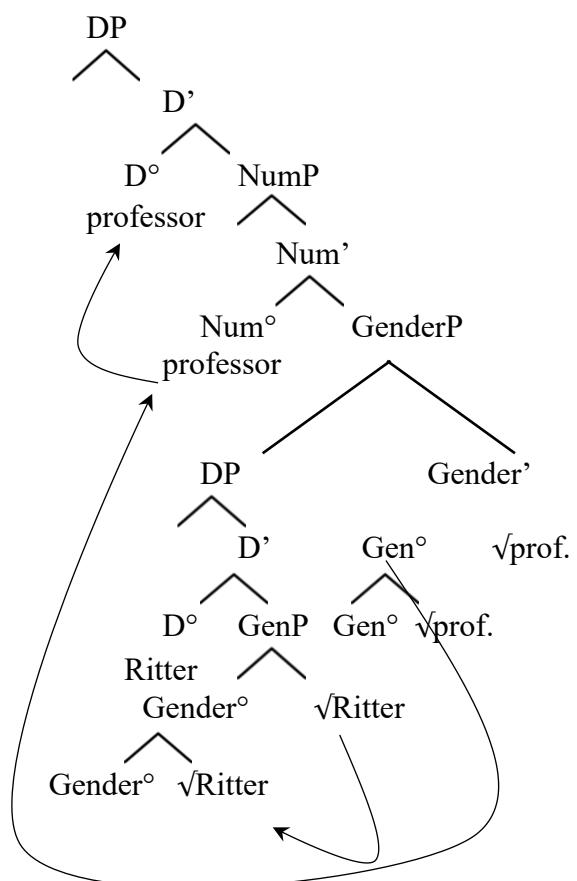
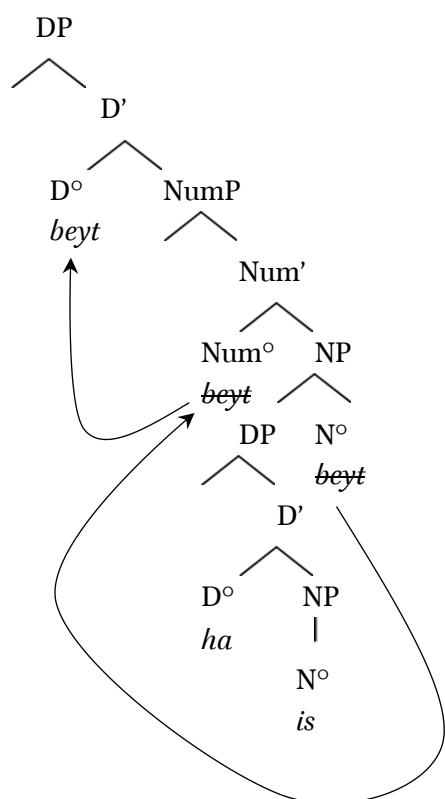
- (90) koningin Ma'thilde [Dutch]  
 queen Mathilde  
 'queen Mathilde'

Note that this stress pattern clearly sets apart the title expression from a compound, as compounding stress would entail stress on the lefthand part in Dutch.

Given the fact that the second part of the construction functions as a proper name, I assume it undergoes N-to-D raising (Longobardi 1994) and as such requires the referential properties of the proper name. Given the empirical similarities of the title expression to the construct state in Ritter (1991), I analyse it further on a par with Ritter's (1991) analysis for the Hebrew construct state given below:

- (91) beyt ha-'is  
housethe-man  
'the man's house'

- (92) professor Ritter  
professor Ritter  
'professor Ritter'



As can be concluded from the structures, the Dutch title expression is analysed as a construct state, but one in which the second part undergoes N-to-D movement, blocking the insertion of a determiner. In other words, it is a referential construct state.

In the construct state, the head noun moves through the nominal inflectional domain. It allows for example, for the possibility of nominal inflection of the head noun in the form of number marking and diminutives:

- (93) professoren Chomsky en Kayne  
professors Chomsky and Kayne  
'professors Chomsky and Kayne'

- (94) prinses-je Elizabeth  
princess-DIMINUTIVE Elizabeth  
'little princess Elizabeth'

- (95) prinses-je-s Elizabeth en Amalia  
princess-DIMINUTIVE-PLURAL Elizabeth and Amalia  
'the little princesses Elizabeth and Amalia'

An accurate depiction of the nominal heads would be far more elaborate than suggested by the tree. I would assume more heads to derive the countability reading, see Borer (2005a) and (De Belder 2011b) for details. NumP is here inserted in the tree as a shorthand for 'heads involved in the countability domain'. On a par with De Belder (to appear, chapter 3) I would also assume a classifier head next to the Gender head.

Note that the agreement in number between the title and the proper name in the construction clearly distinguishes these constructions from compounds. In sum, Dutch has referential construct states alongside referential compounds.

#### 4.6 When is N-to-D movement licit?

In section 4 I noticed a contrast between the referential compound and the -s possessive. It involved the issue that some roots which qualify for unique reference do not undergo N-to-D raising when in argument position. This goes hand in hand with the fact that they cannot occur as the non-head of the s-possessive either:

- (96) # Ik zie zon.  
I see sun

- (97) \* zons zachte warmte  
sun's gentle warmth

Yet, they can be the non-head of the referential compound:

- (98) zonsverduistering  
sun-S-eclipse  
'solar eclipse'

Why are some roots unable to undergo N-to-D raising and what causes the opposition between the referential compound and the -s possessive? In Borer (2005:84-85) the issue is discussed that certain proper names, such as *the Bronx* or *the Pacific Ocean* cannot occur without an article. After some discussion, Borer concludes that 'for reasons we can only speculate on' certain roots are banned from being proper names. She assumes that the reason is to be situated outside of syntax proper.

The present data indeed point in this direction. First, let us assume that for some roots the combination of the definite article plus the root is actually stored at Encyclopedia as the proper way to refer to the entity. For example, *de zon* is the conventional, lexicalised Dutch way to refer to the sun, whereas *Zon* is not. If *zon* 'sun' then would move to D, Encyclopedia would not be able to assign a reference to the construction. More generally, it follows that there is a rather superficial, extra-syntactic ban on the n-to-d movement. Syntax itself does not prohibit the movement. Interestingly, this makes the empirical predictions that if one can alleviate the extra-syntactic limitation, the N-to-D movement should be licit again.

Now consider the syntax of titles again. They do not undergo N-to-D raising in argument position:

- (99) \* Ik feliciteer professor.  
I congratulate professor

- (100) Ik feliciteer de professor.  
I congratulate the professor  
'I congratulate the professor.'

The ban on the N-to-D movement for such roots is rather clear: they lack the typical unique reference which Encyclopedia requires to interpret the structure and there is no idiom stored to interpret the structure either. Note also the default encyclopedic interpretation for N-to-D movement would fail as well: *professor* should not be interpreted as a proper name, the sentence is not about a person whose proper name is Professor. In sum, the structure is uninterpretable.

However, as we have seen, in title expressions the title itself is subject to N-to-D movement:

- (101) Ik feliciteer Hagit met haar verjaardag.  
I congratulate Hagit with her birthday  
'I congratulate Hagit on the occasion of her birthday.'

- (102) Ik feliciteer professor Borer met haar verjaardag.  
I congratulate professor Borer with her birthday  
'I congratulate professor Borer on the occasion of her birthday.'

Crucially, when combined with a proper name, there is no ban on moving to D for a title. After all, why should there be such a ban? *Professor Borer* is fully interpretable: it has unique reference. These observations illustrate that syntax has no general ban for certain roots to move to D, the ban is interpretational.

Consider further the fact that many of the referential compounds under discussion are stored, lexicalised words. In other words, *encyclopedia* has an interpretation stored that matches their structure:

- (103)    zonsverduistering ↔ “SOLAR ECLIPSE”

The fact that the root *zon* ‘sun’ would require the determiner in argument position to refer the star so familiar to us is simply irrelevant for the interpretation of the compound at Encyclopedia. Note that even if the compound had not been stored, the reasoning still holds. The non-head of a compound is a quite unique syntactic position which is arguably subject to its own interpretational rule at Encyclopedia.

It had been noted before that there is no ban on incorporating ‘proper names that require an article’ as in (104) (example taken from Borer 2005a:84, who cites Peter Ackema for suggesting it):

- (104)    Bronx-lover

I conclude that it follows that the set of roots that can occur in r-compounds is a superset of the set that can occur in the -s-possessive: roots with unique reference that would otherwise require a definite article in argument position may occur in the compounds, but not in the -s possessive. More generally, the idea that the ban on N-to-D movement for roots such as *sun* and *Bronx* is extra-syntactic seems to be on the right track.

#### 4.6 Conclusion

In this short contribution I argued that Dutch has referential compounds: compounds with a referential non-head, even though that seems a *contradictio in terminis*. They resemble the Dutch s-possessive in that their non-heads involve n-to-d movement. However, unlike the constructs, their structure contains head incorporation which results in wordhood. The referential compound is therefore an instance of the fourth logical possibility according to the two parameters defined by Borer (2011) for N-N combinations: compounding versus constructs and modificational structures versus referential ones. Together with the syntactic structure of titles plus proper names, the referential compound further contributes evidence to the idea that a ban on N-to-D movement for certain uniquely referring roots, such as *sun* and *Bronx* is extra-syntactic.

## 5 One Module, Different Levels of Merge

### AN(N) compounds in Dutch

Marijke De Belder & Marjo van Koppen

#### Abstract

In this article we provide empirical evidence against the claim that morphology contrasts with syntax in dealing with items that are listed in the lexicon. (Jackendoff 1975, Aronoff 1976, Jensen and Stong-Jensen 1984). More specifically, we distinguish between three types of ANN compounds in Dutch. We show that the structural properties of these types do not show a one-to-one mapping with lexical properties, such as having a listed or even idiomatic meaning (see DiSciullo & Williams 1987). On the basis of this, we argue that conclusions on the structure of certain morphologically complex word-forms should be based on structural properties and not on lexical properties such as idiomaticity or being lexicalized. We propose a syntactic derivation for all types of ANN compounds in Dutch (*pace* Ackema and Neeleman 2004). Structural differences follow from the level of merge: what we traditionally call morphology is syntax below the functional domain.

#### 5.1 Introduction

In this article we distinguish between three types of AN(N) compounds in Dutch (see also Booij 2002, 2010; Ackema and Neeleman 2004; Hüning 2010), which we will refer to as lexicalized compounds, lexicalized phrases and productive compounds. Examples are given in (1)-(3).

(1) lexicalized compounds

- |   |  |
|---|--|
| a. hoog-bouw(-architectuur)<br>high-building(-architecture)<br>'high rise (architecture)' | b. vol-bloed(-paard)<br>full-blood(-horse)<br>'thoroughbred (horse)' |
|---|--|

(2) lexicalized phrases

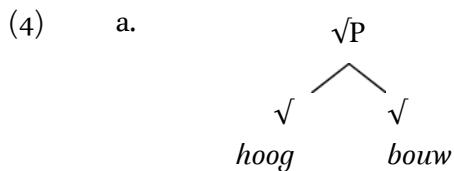
- |   |  |
|---|--|
| a. rode-kool(-recept)<br>red-cabbage(-recipe)<br>'recipe for red cabbage' | b. hoge-school(-student)<br>high-school(-student)<br>'college student' |
|---|--|

(3) productive ANN compounds

- |  |   |
|--|---|
| a. kale-katten-adoptie<br>hairless-cat-adoption<br>'adoption of hairless cats' | b. gezonde-groenten-verkoper<br>healthy-vegetables-seller<br>'seller of healthy vegetables' |
|--|---|

The first two types contain listed or lexicalized material. To be precise, the non-head [AN]-phrase is lexicalized. The third type, on the other hand, is fully transparent and productive in the sense that their creation can remain unnoticed by the native speaker and there is an unbounded possibility to form such compounds (cf. Lieber 1992:3, Schultink 1961 and van Marle 1985:4).<sup>88,89</sup>

Given that word-formation has been associated with being listed (DiSciullo & Williams 1987 *i.a.*), one expects that the productive third type differs dramatically from the other two types that have listed readings. However, it will become clear that this expectation is not borne out. The first type differs considerably from the second type, even though they are both listed. Structurally, the second type is equal to the third type, i.e. the productive compounds. The property of being lexicalized thus crosscuts structural characteristics. We will therefore argue against an association between listedness and structural properties. We will further propose that all three types are derived in syntax by the operation Merge (*pace* Ackema and Neeleman 2004, section 4, but see Booij 2002, 2010 and Hüning 2010 who also assume syntactic structure within these compounds in a Construction Grammar approach).<sup>90</sup> However, the level of Merge may differ, giving rise to structural distinctions. More specifically, we claim that lexicalized compounds are derived via root merger, i.e. merger below category-specific functional projections, as in (4)a. The non-head of the first type is thus a compound consisting of roots that may be embedded in yet another compound, hence the third member is optional. Lexicalized phrases and productive compounds are derived from a partial NP, as in (4)b and (4)c. The non-head of the second and third type is a phrase that may be embedded in a compound, deriving an ANN compound. We will conclude that type 2 and type 3 are structurally similar, as is shown in and (4)b and (4)c. The only difference between the two types is the status in listedness of the left-hand part.

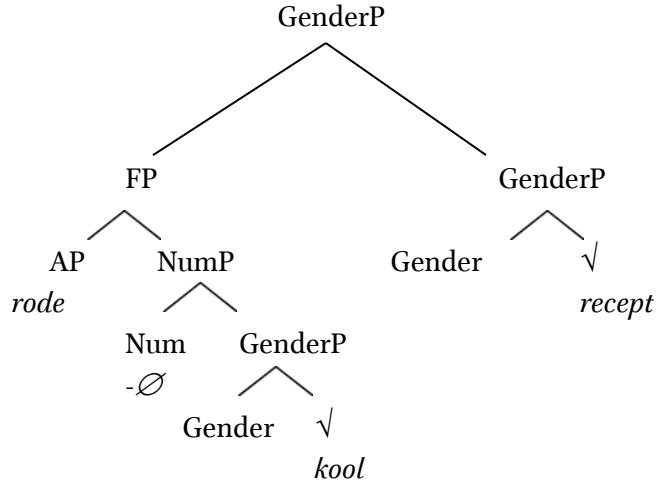


<sup>88</sup> See Hoeksema (1988), De Haas and Trommelen (1993), Booij (2002) and Ackema & Neeleman (2004) for earlier discussions of these data suggesting that [[AN]N] compounds may or may not include a phrase and pointing at the lexicalized nature of some of the data. Ackema & Neeleman (2004) also distinguish between type 1 and 2/3, but they do not distinguish type 2 from type 3. We will come back to Ackema & Neeleman (2004) discussion of these compounds in section 4 below.

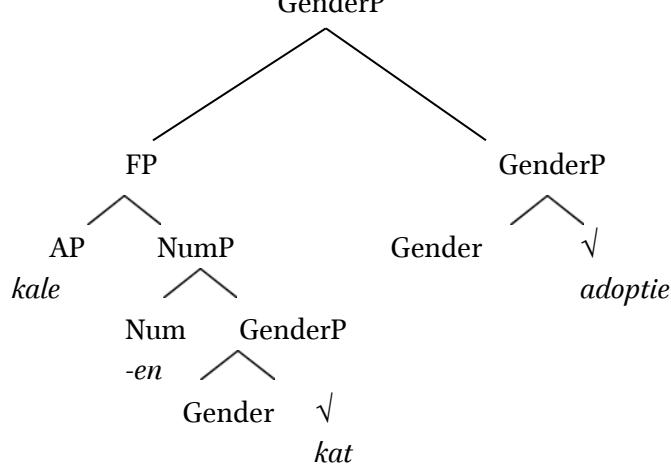
<sup>89</sup> [ANN] compounds are not the only type of compounds in Dutch with a phrasal left-hand part. There are other compounds that might be of the same type, such as CardNN-compounds (e.g. *zeven-mijls-laars* ‘seven-league boot’ (Lit. seven-mile-boot)). Structurally, they may be highly parallel to the ones under discussion in the present article, yet, they are not the focus of this paper. There are further compounds with a quotative left-hand part, such as *blíjf-van-mijn-líjf-huis* ‘women’s shelter’ (Lit. stay-of-my-body-house). These compounds probably differ substantially from the ones discussed in the paper (see Harley 2009 for discussion).

<sup>90</sup> Dutch ANN compounds and AN phrases have also been studied within Construction Grammar (Booij 2002, 2010; Hüning 2010). This work contains observations we do not discuss here, such as observations on the (blocking) relation that exists between AN phrases on the one hand and AN compounds and NN compounds on the other and comparative observations on Dutch and German compounds. Both Booij (2002, 2010) and Hüning (2010) stress the fact that AN compounds indicate that there is no clear schism between syntax and the lexicon. It is claimed that syntactic schemas must exist in the lexicon (Booij 2010: 190). We believe that in spirit, their proposals are similar to what we have in mind: we all emphasise that the compounds contain syntactic structure, even though the way we would model the origin of syntactic structure in these compounds in a very different way, given the different frameworks.

b.



c.



In sum, we will derive structural distinctions from the level of merge. Furthermore, we argue that the level of merge is independent of the property of being lexicalized. Although the first two types contain lexicalized material, only the first type involves root merger. Structural distinctions thus crosscut lexical ones.

To be entirely clear, in the remainder of this paper we will assume the following definition of being lexicalized:

(5) Definition of lexicalized compounds or phrases

Lexicalized compounds or phrases are compounds and phrases that are recognized by the linguistic community as a fixed combination. As a consequence, they can most typically be found in a dictionary.

Note that our definition of being lexicalized does not necessarily involve idiosyncrasy. Although lexicalized compounds or phrases may be idiosyncratic, this is not necessarily the case. For example, the compound in (6)a is lexicalized. It is attested in standard dictionaries of Dutch, such as *Woordenboek der Nederlandsche Taal* or *Van Dale*, which indicates that it is recognized as a listed

word by the speech community. It is indeed a fixed combination, the adjective *hoog* 'high' cannot be replaced by another adjective, as is shown in the *b*- and *c*-examples, even though this is possible in a DP, as in (7).

- (6) a. hoog-conjunctuur  
high-economic climate  
'boom'  
b. \* goed-conjunctuur  
good-economic.climate  
c. \* sterk-conjunctuur  
strong-economic.climate
- (7) een hoge / sterke / goede conjunctuur  
a high / strong / good economic.climate  
'a high/strong/good economic climate'

Despite the fact that this compound is lexicalized, it is fully transparent.<sup>91</sup> There is thus no reason to equate being lexicalized with idiomticity. Note that idiomticity of course implies being lexicalized. If an idiosyncratic meaning is associated with a fixed combination, the combination is necessarily the product of listedness. Idiomticity is thus a stronger notion than being lexicalized. Keep in mind that in this article our claims are not about idiomticity, but about the weaker notion of being lexicalized.

We will proceed as follows. In the following section we introduce the three types of AN(N) compounding in Dutch. We present five characteristics to distinguish between these types. In section 3 we provide an analysis for each type. Section 4 contains a discussion on a previous, contrasting proposal from Ackema and Neeleman 2004. Section 5 sums up and concludes.

## 5.2 Three subtypes of AN(N) compounds in Dutch

### 5.2.1 Introduction

In this section we will present five tests to distinguish between the three types of AN(N) compounds in Dutch. It will become clear that each type is defined by its own characteristics, which include the possibility of degree modification, being lexicalized, the presence or absence of inflection on the adjective, stress patterns and the possibility to contain a comparative or superlative form of the adjective.<sup>92</sup> We then argue that AN(N) compounds containing a comparative or superlative never

---

<sup>91</sup> The meaning is transparent in the sense that there is a clear relation between the meanings of the compound's parts and its meaning as a whole. We do not intend to claim that its actual meaning is the only one which could have been assigned transparently to the compound. As a reviewer points out, it could as well have been the case that *hoogconjunctuur* were listed with a different meaning which is equally transparent, such as 'economic climate at times when the tide is high'.

<sup>92</sup> There is one other property that differentiates between these three types of compounds, namely a restriction on the gender specification of the most deeply embedded noun. Type 3 compounds (see ib), in contrast to type 1 compounds (see ia), resist neuter gender on the left most noun for some speakers and in specific syntactic contexts:

belong to the type of lexicalized compounds. They are invariably instances of either the second type, i.e. the lexicalized phrases, or of the third type, i.e. the productive ANN compounds, depending on their lexicalized status.<sup>93</sup> Finally, we will present examples which seem to suggest there is a yet a fourth type of AN(N) compounding. We will, argue, however, that looks may be deceiving and that this alleged fourth type fully patterns with the first type, i.e. the type of lexicalized compounds, of which it is a subtype. We conclude that Dutch has three types of AN(N) compounds.

### 5.2.2 Five distinguishing characteristics

#### 5.2.2.1 Lexicalization

As the name suggests, lexicalized compounds and lexicalized phrases consist of lexicalized material. In the introduction we have defined being lexicalized as being recognized by the linguistic community as a fixed combination. As a result, such combinations can often be found in dictionaries. This is indeed the case for the examples in (8) and (9), which can be found in *Woordenboek der Nederlandsche Taal* and *Van Dale*.

- (8) hoog-conjunctuur  
high-economic climate  
'boom'

- (9) rode-kool  
red-cabbage  
'red cabbage'

Productive ANN compounds belong to the domain of possible and not necessarily realized words. By definition, they are not listed or lexicalized. It should therefore not come as a surprise that (10), which was made up by the authors of this article, is not a lemma in any dictionary.

- (10) gezonde-groenten-verkoper  
healthy-vegetables-seller  
'seller of healthy vegetables'

Other neologisms that we have found via Google are given in (11).

- 
- |        |   |    |   |   |
|--------|---|----|---|---|
| (i) a. | hoog-seizoen-tarief<br>high-season <sub>neuter</sub> -price<br>'price in the high season' | b. | * | sterk-ijzer-verkoper<br>strong-iron <sub>neuter</sub> -seller<br>intended: seller of strong iron' |
|--------|---|----|---|---|

The restriction on (i)b results from the fact that for type 3-compounds the adjectival inflection needs to be computed. However, Dutch neuter adjectival inflection is dependent on the definiteness of the DP. Since the left-hand part of the compound systematically lacks a D-layer (see footnote 104) the computation of the adjectival inflection is impossible. Hence, the ungrammaticality of type 3 compounds with a neuter left-hand noun.

<sup>93</sup> We will conclude that type 2 and 3 are structurally similar. For this reason one might argue that they are two subtypes of a single compounding type.

- (11) a. oude-schoenen-actie<sup>94</sup>  
          old-shoe-action  
          'sales action involving old shoes'
- b. kleine-mensen-dating<sup>95</sup>  
          little-human.beings-dating  
          'dating for small people'

Whether a word is attested in a dictionary is a first indication of listedness, but it may be subject to happenstance. It is therefore noteworthy that two other independent properties correlate with being lexicalized. Firstly, we have pointed out in the introduction that parts of a lexicalized expression cannot be replaced by synonyms. This was illustrated for *hoogconjunctuur* 'boom'. The relevant example is repeated below.

- (12) \* goed-/sterk-conjunctuur  
          good-/strong-economic.climate

In lexicalized phrases substitution by a synonym is penalized by a loss of the lexicalized meaning. The compounds in (13), for example, do not refer to a red cabbage.

- (13) bordeaux- / scharlaken- / purper-kool  
          burgundy- / scarlet- / purple- cabbage  
          'cabbage which happens to be burgundy / scarlet red / purple of color'

A similar substitution is harmless, though, for productive ANN compounds. Given that their meaning is not lexicalized, no lexicalized meaning can be lost. The compositional meaning is thus retained if a synonym of the adjective is used, as (14) illustrates.

- (14) heilzame-groenten-verkoper  
          wholesome-vegetables-seller  
          'seller of wholesome vegetables'

In sum, lexicalized combinations do not allow a substitution of its parts by a synonym, whereas non-lexicalized productive compounds do allow such a replacement.

A second property which correlates with being lexicalized is the possibility of a meaningful stacking of the adjective or its antonym. In lexicalized compounds and phrases such a stacking is interpretable. This can be seen in the examples below. The *a*-examples show that the same adjective can be stacked, the *b*-examples show a similar effect with the adjective's antonym.

- (15) a. Peking heeft hoge hoog-bouw .  
          Beijing has high high-rise  
          'Beijing has high high-rise'.

---

<sup>94</sup> <http://www.brantano.be/nl/oude-schoenenactie/hoe-de-bonnen-gebruiken> (May 15, 2014)

<sup>95</sup> <http://www.kleinemensen-dating.nl/Dating/Dating.aspx> (May 15, 2014)

- b. De Amsterdamse Zuidas heeft **lage hoog-bouw**.  
 The Amsterdam South.axis has low high-rise  
 'The Amsterdam South Axis has low high-rise'.

- (16) a. Een lage pH-waarde geeft **rode rode-kool**.  
 a low pH-value yields red red-cabbage  
 'A low pH-value yields red red cabbage'.  
 b. Een hoge pH-waarde geeft **blauwe rode-kool**.  
 a high pH-value yields blue red-cabbage  
 'A high pH-value yields blue red cabbage'.

A comparable combination of adjectives leads to sheer nonsense in productive ANN compounds, which is marked by means of a hashmark.

- (17) a. # gezonde-gezonde-groenten-verkoper  
 healthy-healthy-vegetables-seller  
 b. # ongezonde-gezonde-groenten-verkoper  
 unhealthy-healthy-vegetables-seller

To conclude, in this subsection we have emphasized on a lexical distinction between lexicalized compounds and lexicalized phrases on the one hand and productive ANN compounds on the other. The first two types are lexicalized, i.e. they are recognized by the speaker as a fixed combination. As a result, it is not possible to substitute their parts by synonyms. However, it is possible to stack the adjective or its antonym in a meaningful way. The third type differs in this respect. It does not contain fixed combinations of vocabulary items. Consequently, it tolerates the substitution of its parts by synonyms. In contrast, a stacking of the adjective with the same adjective or its antonym leads to jabberwocky.

#### 5.2.2.2 Degree modification

In lexicalized compounds the adjective cannot be modified by a degree modifier. It is thereby irrelevant whether the adverb of degree is understood as embedded within the compound or as a modifier of the entire compound. This is illustrated in (18).

- (18) a. \* [erg-hoog]-conjunctuur  
 very-high-economic.climate  
 b. \* erg [hoog-conjunctuur]  
 very high-economic.climate

Similarly, lexicalized phrases cannot contain a degree modifier either, as shown in (19).

- (19) \* erg-rode-kool-recept  
 very-red-cabbage-recipe

This restriction seems to be connected to the fact that the AN phrase contained within the compound is lexicalized. If one combines a degree modifier with the lexicalized phrase itself, the lexicalized meaning is lost and only a productive, *ad hoc* meaning can be assigned to the phrase.

- (20) een erg rode kool  
       a very red cabbage  
       'a cabbage which happens to be very red'

In contrast, productive ANN compounds containing a degree modifier are grammatical, as can be deduced from (21).

- (21) erg-gezonde-groenten-verkoper  
       very-healthy-vegetables-seller  
       'seller of very healthy vegetables'

In sum, lexicalized compounds and lexicalized phrases cannot contain an adverbial degree modifier, whereas productive ANN compounds do tolerate such a modifier.

### 5.2.2.3 Inflection on the adjective

In this section we will discuss whether the adjective which is contained in the AN(N)-compound bears adjectival inflection. It will become clear that this is not the case for lexicalized compounds, whereas adjectival inflection is attested in lexicalized phrases and productive ANN compounds.

Lexicalized compounds do not contain adjectival inflection. This becomes clear when we compare the compound in (22)a with the indefinite and definite DPs in (23). Both DPs show a schwa ending on the adjective which realizes inflection. Such a schwa is obligatorily absent in the compound, as the *b*-example in (22) illustrates.

- (22) a. hoog-conjunctuur  
          high-economic.climate  
          'boom'  
       b. \* hoog-e-conjunctuur  
          high-INFL-economic.climate
- (23) a. een hoog-e conjunctuur  
          a high-INFL economic.climate  
          'a strong economic climate'  
       b. de hoog-e conjunctuur  
          the high-INFL economic.climate  
          'the strong economic climate'

In this respect they differ from lexicalized phrases or productive ANN compounds which do contain inflection on the adjective. Let us first consider lexicalized phrases. As can be seen in (24), the adjective is marked with an inflectional ending, on a par with the adjective in the DPs in (25).

- (24) rood-e-kool  
 red-INFL<sup>96</sup>-cabbage  
 'red cabbage'
- (25) a. een rood-e kool  
 a red-INFL cabbage  
 'a cabbage which happens to be red'
- b. de rood-e kool  
 the red-INFL cabbage  
 'the cabbage which happens to be red'

Adjectival inflection is attested in productive ANN compounds as well as can be deduced from a comparison between the compound in (26) and the DPs in (27).

- (26) gezond-e-groenten-verkoper  
 healthy-INFL-vegetables-seller  
 'seller of healthy vegetables'
- (27) a. gezond-e groenten  
 healthy-INFL vegetables  
 'healthy vegetables'
- b. de gezonde groenten  
 the healthy-INFL vegetables  
 'the healthy vegetables'

In sum, adjectival inflection differentiates between lexicalized compounds on the one hand and lexicalized phrases and productive ANN compounds on the other hand. While lexicalized compounds lack adjectival inflection, it is attested in the other two types.

#### 5.2.2.4 Stress

DP stress differs from compound stress in Dutch. DPs bear stress on the noun, as in (28), compounds bear stress on the leftmost part, as in (29). Small caps indicate stress.

- (28) een slimme VROUW  
 a smart woman  
 'a smart woman'

---

<sup>96</sup> De *-e-* is adjectival inflection. It is most definitely not a so-called linking element. First of all, linking elements have a different shape, they are either *-s-* or *-en-* (the /n/ in *-en-* is audible in certain dialects (Hanssen 2012)). Secondly, the adjectival inflection is sensitive to the gender of the left-hand N in these compounds (see footnote 92). Thirdly, it appears in regular syntactic phrases as well, as in (25)b. Needless to say, there is no reason why a linking element should occur in a syntactic phrase.

- (29) een TONG-zoen  
 a tongue-kiss  
 'a French kiss'

The question is now whether ANN compounds bear DP stress or compound stress. Lexicalized compounds typically get compound stress. This is shown in (30).

- (30) a. HOOG-bouw(-architect)  
 high-building(-architect)  
 'high rise architect'
- b. ZOET-hout-verkoper  
 sweet-wood(-seller)  
 '(seller of) liquorice'
- c. VOL-bloed(-kwekerij)  
 full-blood(-farm)  
 'farm producing thorough-bred animals'
- d. BITTER-koekjes(-pudding)  
 bitter-cookies(-pudding)  
 'macaroons pudding'

In the examples above stress is assigned to the leftmost part, clearly indicating these compounds bear compound stress.<sup>97</sup> Admittedly, though, some idiolectal variation is attested for a minority of these compounds (see Haeseryn et al. 1997: section 12.3.2.4 iii). For example, both (31)a and (31)b are attested.

- (31) a. KLEIN-geld(-portemonnee)  
 small-money(-purse)  
 '(purse for) coins'
- b. klein-GELD(-portemonnee)  
 small-money(-purse)  
 '(purse for) coins'

Stress patterns may thus vary marginally. Nevertheless, it is clear that lexicalized compounds most often take regular compounds stress.

Lexicalized phrases are assigned DP stress, as in the examples below. The left-hand AN phrase which is embedded in the compound is assigned stress as if it were an independent DP, compare (32) a and b. Some more examples are given in (33).

---

<sup>97</sup> Note that example (30)d contains a trochaic adjective. As such, it has the same syllable structure as a monosyllabic adjective with inflection. The fact that the stress in this example is still on the adjective indicates that the stress assignment is dependent on the structure of the compound, rather than on the syllable structure of the adjective.

- (32) a. rode-KOOL(-recept)  
red-cabbage(-recipe)  
'(recipe for) red cabbage'

- b. de rode KOOL  
the red cabbage  
'the red cabbage'

- (33) a. blinde-VINK(-recept)  
blind-finches(-recipe)  
'(recipe for) beef olives'

- b. dikke-DARM(-ontsteking)  
thick-intestine(-inflammation)  
'(inflammation of the) large intestine'

Again, one may find some exceptional idiolectal variation in this domain. For example, (34)a and (34)b are both attested.

- (34) a. hoge-SCHOOL(-student)  
high-school(-student)  
'college (student)'

- b. HOGE-school(-student)  
high-school(-student)  
'college (student)'

In spite of such marginal counter-examples, it is still fair to state that the regular stress pattern for lexicalized phrases is DP stress.

DP stress is also attested in productive ANN compounds, as shown in the examples below.

- (35) a. gezonde-GROENTEN-verkoper  
healthy-vegetables-seller  
'seller of healthy vegetables'

- b. kale-KATTEN-adoptie  
hairless-cats-adoption  
'adoption of hairless cats'

- c. slimme-VROUWEN-vereniging  
smart-women-association  
'association for smart women'

To conclude, both compound stress and DP stress is attested amongst ANN compounds. Lexicalized compounds take compound stress, whereas lexicalized phrases and productive ANN compounds take DP stress.

#### 5.2.2.5 Comparatives and superlatives in ANN compounds

ANN compounds in Dutch may contain the comparative or superlative form of an adjective. Examples are shown below.

- (36) a. hoger-huis-lid  
higher-house-member  
'member of the House of Lords'
- b. hogere-machts-vergelijking  
higher-power-equation  
'equation of higher degree'

The compounds containing a comparative or superlative may be lexicalized; they are recognized as fixed combinations and they loose their meaning if the adjective is replaced by means of a synonym, as shown in (37).

- (37) a. verheven-huis-lid  
elevated-house-member  
'member of an elevated house'
- b. verheven-machts-vergelijking  
elevated-power-equation  
'equation of an elevated degree'

However, non-lexicalized, newly formed examples are not excluded either. Examples (38)a and (38)b are neologisms.

- (38) a. veiligere-narcose-ontwikkeling  
safer-anesthesia-development  
'development of safer anesthesia'
- b. oudste-dochter-verantwoordelijkheid  
oldest-daughter-responsibility  
'responsibility of the oldest daughter'<sup>98</sup>

---

<sup>98</sup> Note that the leftmost noun in these examples with type 3 compounds cannot be neuter singular because of an additional independent restriction on this type of compounds, see footnote 92. There is one exception on this restriction that becomes important here: if the adjective cannot show inflection due to its morphological shape, like *verheven* 'elevated' in ((37)a, neuter singular nouns (like *huis* 'house' in this example) are allowed to occur.

Whether the example is lexicalized or not, we argue that the comparative or superlative contains adjectival inflection. Compare the compounds in (36) with the DPs in (39) and (40). (39) shows indefinite DPs, (40) shows definite DPs.

- (39) a. een hoger-∅ huis<sub>NEUTER</sub>  
a higher-INFL house  
'a higher house'
  - b. een hoger-e vergelijking<sub>COMMON</sub>  
a higher-INFL equation  
'a higher equation'
- (40) a. het<sub>NEUTER</sub> hoger-e huis<sub>NEUTER</sub>  
the higher-INFL house  
'the higher house'
  - b. de<sub>COMMON</sub> hoger-e vergelijking<sub>COMMON</sub>  
the higher-INFL equation  
'the higher equation'

The adjectives in the compounds in (36) select a null inflectional marking if the noun they agree with is marked for neuter gender, see (36)a., they take a schwa ending otherwise, see (36)b. As such, they pattern with the inflection paradigm of indefinite DPs, which selects a null morpheme if the noun is neuter and singular and a schwa in all other cases. We take this to mean that the adjectives in (36) show adjectival inflection.

The fact that ANN compounds containing a comparative or superlative take adjectival inflection, indicates that they never belong to the first type, i.e. the lexicalized compounds. They either belong to the second type, i.e. the lexicalized phrases, or to the third type, i.e. the productive ANN compounds. We therefore propose that the lexicalized examples are instances of lexicalized phrases, whereas the neologisms are examples of productive ANN compounds.

This property of containing a comparative or superlative, relates to the ability to have degree modification (see section 2.2.2), since the comparative morpheme is a degree modifier (see Kennedy 1997). This is indeed what we find for productive ANN compounds. However, it seems to suggest that the lexicalized ANN compounds with a comparative or superlative are wrongly classified as lexicalized phrases as we have suggested that lexicalized phrases do not select degree modifiers (see section 2.2.2). Recall, though, that the two lexicalized types, i.e. the lexicalized compounds and the lexicalized phrases, cannot select degree modifiers as they loose their idiomatic meaning if they do. The relevant examples are repeated in (41) and (42).

- (41) a. \* [erg-hoog]-conjunctuur  
very-high-economic.climate
- b. \* erg [hoog-conjunctuur]  
very high-economic.climate

- (42) \* erg-rode-kool-recept  
           very-red-cabbage-recipe

One may interpret these examples in two ways. One may conclude that they show that lexicalized types cannot select a degree modifier, as we have done above. Alternatively, one may conclude that these examples show that the degree modifier should be part of the lexicalized information. If we follow the second line of reasoning the ANN compounds containing a comparative may still pattern with the lexicalized types although they contain a degree modifier, since the degree morpheme is simply part of the listed information. We then predict that the examples loose their idiomatic meaning in the absence of a degree modifier. This prediction is borne out, as is shown in (43).

- (43) a. hoog-huis-lid  
           high-house-member  
           'member of a high house'
- b. hoge-machten-vergelijking  
           high-power-equation  
           'equation of a high power'

We therefore modify the conclusion that lexicalized types cannot contain a degree modifier. More accurately stated, it is illicit to add or remove a degree modifier from a listed combination. The degree modifier thus suggests that these examples indeed may belong to the type of lexicalized phrases.

Finally, we predict that if these compounds are classified as productive ANN compounds or lexicalized phrases, they take DP stress.<sup>99</sup> Now observe that Dutch DPs containing a comparative or superlative assign stress to the comparative, as in (44). Small caps indicate stress.

- (44) een SLIMMERE vrouw  
       a      smarter     woman  
       'a smarter woman'

- (45) de SLIMSTE vrouw  
       the    smartest woman  
       'the smartest woman'

Recall that compound stress is assigned to the leftmost member of the compound, as in (46).

- (46) een TONG-zoen  
       a      tongue-kiss  
       'a French kiss'

---

<sup>99</sup> A reviewer points out that s/he gets exceptional stress assignment in the DP *de hogere MACHT* 'the higher power'. However, s/he admits that the same stress patterns pertains to the compound. This reinforces the claim that the stress in the compound systematically mimics the stress in the DP.

The leftmost member of an ANN compound containing a comparative or superlative is of course the comparative or superlative adjective itself. As a consequence, we expect stress to fall on the comparative or superlative both in the case of DP stress and in the case of compound stress, rendering the test inconclusive. The comparative or superlative will be assigned stress according to either stress pattern. Unsurprisingly, it is indeed the comparative or superlative adjective which is assigned stress.

- (47) a. HOGER-huis-lid  
higher-house-member  
'member of the House of Lords'
- b. HOGERE-machts-vergelijking  
higher-power-equation  
'equation of higher degree'

Although the data thus fully behave as expected, the test is inconclusive for this type of ANN compounding. We conclude that the data are in any case compatible with the conclusion reached above, i.e. that ANN compounds containing a comparative or superlative are always examples of lexicalized phrases or productive ANN compounds and never of lexicalized compounds.

In short, in this section we have shown that ANN compounds containing a comparative or superlative pattern with lexicalized phrases and productive ANN compounds, depending on their status as being lexicalized or newly formed. They are never lexicalized compounds. The main indication leading to this conclusion is the fact that they take adjectival inflection. We have further pointed out that the degree modifier, i.e. the comparative or superlative morpheme, can be part of the lexicalized phrase. We have therefore modified the previous conclusion that lexicalized phrases cannot contain a degree modifier. What is rather at play is that one should not tinker with the stored information on the presence or absence of degree modifiers if one wants to retain the lexicalized meaning. Finally, we have pointed out that a test based on stress patterns is inconclusive in these cases as compounds stress and DP stress both happen to assign stress to the comparative or superlative adjective. We have concluded that at least the stress pattern does not contradict the conclusion we have reached above that ANN compounds containing a comparative or superlative are instances of lexicalized phrases or productive ANN compounds.

#### 5.2.2.7 A fourth type?

There is yet another set of AN(N) compounds we have not discussed so far. These compounds typically contain an adjective which is followed by a schwa, as in (48).

- |   |   |
|---|---|
| (48) a. wit-e-brood<br>white-E-bread<br>'white bread' | b. zoet-e-melk<br>sweet-E-milk<br>'plain milk' <sup>100</sup> |
| c. plat-e-land<br>flat-E-land<br>'countryside'        | d. vast-e-land<br>fixed-E-land<br>'mainland'                  |

---

<sup>100</sup> *Zoetemelk* is the antonym of buttermilk.

- |   |  |
|---|--|
| e. blind-e-man<br>blind-E-man<br>'blindfolded player playing tag' | f. wild-e-bras<br>wild-E-brute<br>'tear-away'      |
| g. mal-e-molen<br>crazy-E-mill<br>'carousel'                      | h. mal-e-praat<br>crazy-E-talk<br>'drivel'         |
| i. dol-e-praat<br>mad-E-talk<br>'drivel'                          | j. hard-e-bol<br>tough-E-ball<br>'stubborn person' |
| k. oud-e-jaar<br>old-E-year<br>'New Year's Eve'                   | l. wild-e-man<br>wild-E-man<br>'tear-away'         |

In this section we will argue that these compounds are a subtype of the first type, i.e. of the lexicalized compounds. We will therefore refer to them as lexicalized compounds with an intervening schwa. We thus expect them to disallow degree modification, to be lexicalized, to lack adjectival inflection and to be assigned compound stress. These expectations are borne out.

First consider degree modification. Example (49) shows that this type of compounding indeed loses its specific meaning when a degree modifier is added.

- (49) een erg wilde-man  
 a very wild man  
 'a very wild man'  
 \* 'a terrible tear-away'

From the fact that most examples in (48) are clearly idiomatic one can deduce that they are lexicalized. After all, idiomticity is a stronger notion than being lexicalized. If a compound is idiomatic this implies that the speech community attaches a non-compositional meaning to a fixed combination. It follows that the combination is indeed fixed, i.e. lexicalized.

We further predict that these compounds lack adjectival inflection. This is indeed the case, although they contain a schwa which at first sight could be mistaken for adjectival inflection. We have seen in section 2.2.6 that if the modified noun in a compound is a neuter noun, adjectival inflection is marked by means of a null morpheme, the schwa being restricted to common nouns. If the schwa truly were adjectival inflection, we expect it to be selected uniquely by compounds with a common noun. However, it co-occurs with neuter nouns as well, as can be deduced from the examples below. The examples in (50) show that the nouns *brood* 'bread' and *land* 'land' are indeed neuter. The examples in (51) show that the gender of these nouns does not affect the presence of the intervening schwa.

- (50) a. het<sub>NEUTER</sub> brood<sub>NEUTER</sub>  
          the       bread  
          'the bread'
- b. het<sub>NEUTER</sub> land<sub>NEUTER</sub>  
          the       land  
          'the land'

- (51) a. wit-e-brood  
          white-E-bread  
          'white bread'
- b. plat-e-land  
          flat-E-land  
          'country-side'

From the fact that the intervening schwa is not blocked by neuter nouns, we conclude it is not adjectival inflection. One may then wonder what the status of the schwa is. We conjecture it is there for phonological reasons. Considering the examples in (48) it is noticeable that the adjectives invariably end in a dental or a lateral consonant. Given the fact that one can attest such a phonological pattern, it is not unreasonable to assume the phonological context triggers the schwa, albeit for reasons we fail to understand. This hypothesis is further supported by the contrasting pair in (52).

- (52) a. oud-e-jaar  
          old-E-year  
          'New Year's Eve'
- b. nieuw-(\*)e-jaar  
          new-year  
          'New Year'

There is no obvious reason to expect a structural difference between (52)a and (52)b. Yet, (52)a contains a schwa, whereas (52)b does not. This patterns with a phonological distinction; (52)a ends in a dental, whereas (52)b does not. In short, we conclude that the intervening schwa is triggered phonologically. It is not a realization of adjectival inflection. Note that the insertion of the schwa must have been but an optional phonological rule as a form without an intervening schwa may exist as well. For example, (53) has been stored alongside (52)a.

- (53) oud-jaar  
          old-year  
          'New Year's Eve'

Finally, we expect these examples to show compound stress. This expectation is indeed borne out, as is illustrated in (54).

- (54) a. WIT-e-brood  
          white-E-bread  
          'white bread'
- b. ZOET-e-melk  
          sweet-E-milk  
          'plain milk'

We have seen in section 2.2.4 that some examples of lexicalized compounds may show a deviating stress pattern. Admittedly, this is also the case for the examples under discussion. Both (55)a and (55)b are attested.

- (55) a. OUD-e-jaar  
          old-E-year  
          'New Year's Eve'
- b. oud-e-JAAR  
          old-E-year  
          'New Year's Eve'

In short, the stress pattern of these compounds is fully compatible with the hypothesis that they are a subtype of lexicalized compounds. To summarize, from the fact that these compounds show all the characteristics of lexicalized compounds, we propose they are a subtype thereof. The intervening schwa is triggered by the phonological context.

### 5.2.3 Conclusion

We have presented four tests to differentiate between three types of ANN compounds in Dutch. These tests involve distinctions in being lexicalized, in the possibility to host an adverbial degree modifier, in allowing adjectival inflection, in stress patterns and in allowing a comparative or a superlative. Finally, we have shown that lexicalized compounds may contain an intervening schwa which is triggered by the phonological make-up of the adjective. The results of this section are summarized in the table in (56).

(56)

	Lexicalized compounds with or without an intervening schwa	Lexicalized phrases <sup>101</sup>	Productive ANN compounds
lexicalized	✓	✓	✗
adding degree modification	✗	✗	✓
adjectival inflection	✗	✓	✓
DP stress	✗	✓	✓
comparative or superlative	✗	✓	✓

Overview of ANN compounds in Dutch

A closer look at this table reveals that lexicalized compounds and lexical phrases pattern alike when it comes to lexical properties, such as being lexicalized, and that lexical phrases and productive ANN compounds are similar structurally. Lexical properties thus cross cut structural ones. In the next section we develop this observation in full detail.

## 5.3. Analysis: Root merger vs. NP merger

### 5.3.1 Introduction

In this section we argue that lexicalized compounds involve bare root merger, whereas the left-hand part of lexicalized compounds and productive ANN compounds is truly phrasal, i.e. it contains a partial NP. We first discuss lexicalized compounds and then we focus on productive ANN compounds. Lexicalized phrases are analyzed last.

---

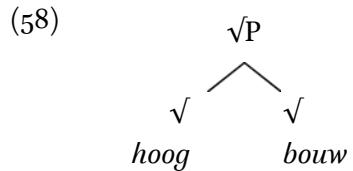
<sup>101</sup> The term 'lexicalized phrases' is short for 'compounds with a lexicalized phrase as their left hand part'.

### 5.3.2 Root merger of lexicalized compounds

In this section we will present an analysis of lexicalized compounds, as in (57).

- (57)      a. hoog-bouw(-architectuur)  
               high-building(-architecture)  
               'high rise (architecture)'  
  
       b. zoet-hout(-verkoper)  
               sweet-wood-seller  
               '(seller of) licorice'  
  
       c. klein-kunst(-theater)  
               small-art-theatre  
               'cabaret (theatre)'

We adopt the view that the core of a lexical projection is a categoriless root (Halle and Marantz 1993, Harley and Noyer 1999, Borer 2005a). We propose that lexicalized compounds involve the merger of bare roots, as in (58).<sup>102, 103</sup>



In the structure in (58) the root which we referred to as an adjective is not an adjective in the technical sense of the word. It is but a bare, a-categorial root.

The following empirical properties follow from this structure. It is expected that lexicalized compounds do not take adjectival inflection as there is no adjectival structure present in the tree. In the same vein, degree modifiers cannot merge as there is no AP to host them.

Note that the structure in (58) is derived in syntax and not in a separate morphological module. The reader may wonder how it follows that lexicalized compounds are invariably lexicalized. Although they are productive, they cannot be formed on the spot (Ackema & Neeleman 2004).

- (59)      a. \* fris-wind  
               fresh-wind  
       b. \* blauw-oog  
               blue-eye

---

<sup>102</sup> The structure is simplified for ease of exposition. A fully developed derivation of compounds would take us too far afield. We would like to refer the reader to Borer (2009, 2013b) for a detailed derivation of compounds in a root-based framework.

<sup>103</sup> For a detailed discussion on the merger, projection and linearization properties of bare roots, see De Belder (2011a) and chapter 2.

Despite the productivity, the licitness of the newly formed compound depends on listedness. The speech community has to recognize the new compound as a (newly) listed combination. We will address this issue in what follows.

We adopt the view that bare roots are categorized by means of the functional projections which merge on top of it. For example, if number marking and a D-layer merge above a root, the root becomes nominal, if tense merges above a root, it will be verbal. Categorization is thus not done by categorial heads or lexical specifications (Borer 2005a, 2013a, De Belder 2011a). As a consequence, adjectives are defined by designated functional projections as well, which we assume to be projections of degree (Corver 1990, 1997, Kennedy 1999). Technically, an adjective is thus not a particular lexical projection. It is rather defined as a relation between a predicate as expressed by a root and a degree established by means of functional projections. Roots thus depend on projections of degree to function as adjectives. To compute the meaning of an adjective in a compositional fashion, LF thus depends on degree projections as well. Now observe that bare roots as in (58) lack such functional projections. Hence, an adjectival meaning cannot be computed compositionally. The only possible meaning for the structure in (58) is therefore a stored one, which is not necessarily idiomatic (see section 1). As such, we derive the connection between being lexicalized and root merger.

To summarize, we have analyzed lexical compounds as instances of root merger. We have shown that their empirical properties can be derived from this structure. For example, it is expected that they resist adjectival properties such as degree modification and adjectival inflection. Finally, being lexicalized results immediately from the defective structure.

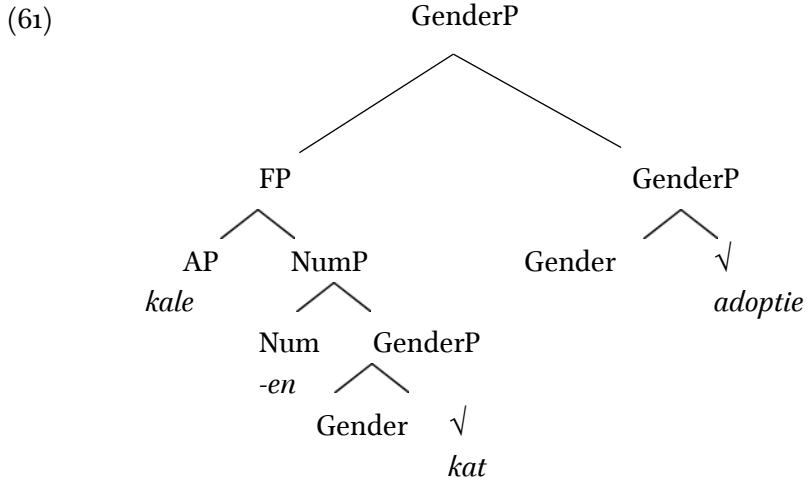
### 5.3.3 NP merger of productive ANN compounds

We propose that the non-head of a productive ANN compound, as in (60), is a partial NP<sup>104</sup>, as in (61). Syntactically, it is thus a phrase.

- |   |   |
|---|---|
| (60) a. kale-katten-adoptie<br>hairless-cat-adoption<br>'adoption of hairless cats' | b. gezonde-groenten-verkoper<br>healthy-vegetables-seller<br>'seller of healthy vegetables' |
|---|---|

---

<sup>104</sup> The non-head of a compound cannot be referential (Hoeksema 1988). This correlates with the fact that it can never include a D-layer (see Harley 2009, Borer 2011). One can find examples with a possessive pronoun, though, such as *Onzelievevrouwewerk* (our-sweet-lady-church) 'Church of Our Lady', in which case the possessive pronoun is part of a proper name. The fact that it does not function as a proper possessive pronoun in these examples can be deduced from the fact that it is fixed, hence it cannot be replaced as in \**jullielievevrouwewerk* (your-sweet-lady-church).



The presence of an AP-layer in (61) captures the fact that productive ANN compounds may contain adjectival features. We have pointed out that these compounds allow degree modifiers and comparative and superlative morphemes and that they contain adjectival inflection. These properties follow immediately from the structure. After all, it is expected that an AP phrase can contain material which is typically associated with such a phrase. Furthermore, we see no reason why this structure should not be productive; NPs can be built as you like. Finally, note that one predicts that the adjective can be coordinated with another one or that several adjectives can be stacked. This is borne out. (62) shows an example of coordinated adjectives, (63) illustrates stacking.

- (62) kale en zieke-katten-adoptie  
 hairless and ill-cats-adoption  
 'adoption of hairless and ill cats'

- (63) kale-zieke-katten-adoptie  
 hairless-ill-cats-adoption  
 'adoption of hairless, ill cats'

In sum, in order to capture the fact that the adjective in productive ANN compounds shows the syntactic behavior of a full-blown adjective, we propose to analyze the left-hand AN phrase of the compound as a partial NP with a complete AP layer.

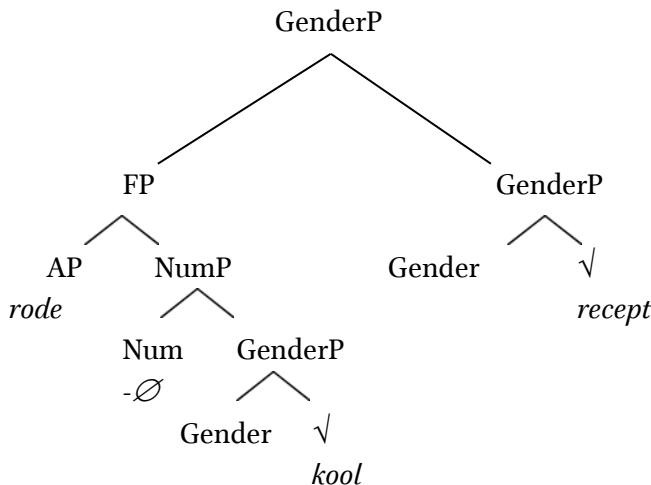
### 5.3.4 NP merger of lexicalized phrases within compounds

Lexicalized phrases which are embedded in a compound share structural properties with productive ANN compounds. Both types contain adjectival inflection and both may contain a comparative or superlative form of the adjective. We therefore propose to analyze lexicalized phrases on a par with productive ANN compounds. In other words, lexicalized phrases within compounds are partial NPs as well, as in (64).<sup>105</sup>

---

<sup>105</sup> We propose a number phrase as the left-hand phrase can be pluralized, as in *rodekolenveld* 'red cabbage field'. We propose a gender phrase as the adjectival inflection is clearly sensitive to gender, as can be concluded from the zero adjectival marker which is triggered by the neuter diminutive affix in *het rood-Økooltje* 'the small red cabbage'.

(64)



However, we have discussed a distinction between productive ANN compounds and lexicalized phrases. Whereas one can add a degree modifier to productive ANN compounds, this option is excluded for lexicalized phrases. Yet, we have argued that this difference does not stem from the fact that lexicalized phrases resist a degree modifier. After all, they may contain a comparative or superlative morpheme. What is relevant, though, is the fact that lexicalized phrases can only contain degree modification if it is part of the stored information. This distinction between productive ANN compounds and lexicalized phrases is thus not a structural distinction. It rather stems from the lexical property of being lexicalized.

A similar effect can be observed when considering the coordination of adjectives in lexicalized phrases. They do allow coordination with other adjectives, as expected. This is shown in (65).

- (65) a. nieuw-bouw  
new-build  
'new build'
- b. hoog-bouw  
high-build  
'high rise'
- c. nieuw- en hoog-bouw<sup>106</sup>  
new- and high build  
'new high rise'

As lexicalized phrases contain an AP-layer they thus can be coordinated with other adjectives. Yet, they cannot be coordinated with just any adjective, as is illustrated in (66).

---

<sup>106</sup> It is easy to find examples on Google in which the compound *nieuw- en hoogbouw* unambiguously refers to the same building, e.g. *De nieuw- en hoogbouw aan de Laan op Zuid vordert met de dag. Het zal niet lang meer duren of de Vancouver Building is klaar.* (<http://renehoeblaak.com/2009/12/>) ‘The new high rise at the South Avenue progresses daily. It won’t take long till the Vancouver Building is completed.’

- (66) a. mooie bouw  
 nice build  
 'nice build'
- b. hoog-bouw  
 high-build  
 'high rise'
- c. \* mooie- en hoog-bouw  
 nice- and high-build

We propose that the illicitness of (66)c is not due to a structural restriction, it rather goes back to a lexical restriction. Note that the DP in (66)a is fully compositional, it is not lexicalized. Now observe that although listed items can be coordinated, this is not necessarily the case for listed and non-listed items. Example (67) shows examples of listed imprecations.

- (67) a. Krijg de tyfus.  
 get the typhoid  
 (imprecation)
- b. Krijg de tering  
 get the tuberculosis  
 (imprecation)

Although the imprecation may contain various names of diseases, not all diseases can be part of the idiom. Example (68), for example, is not a listed imprecation.

- (68) Krijg de griep.  
 get the flu  
 'Get the flu.'<sup>107</sup>  
 \* imprecation

One can licitly coordinate the names of the diseases as long as they are both listed as imprecations, as can be seen in (69). However, coordinating listed and non-listed items is ungrammatical, as is shown in (70).

- (69) Krijg de tyfus en de tering.  
 get the typhoid and the tuberculosis  
 (imprecation)

---

<sup>107</sup> This example is pragmatically odd as it is unlikely one would order someone to catch a disease. An interpretation as an imprecation is excluded.

- (70) \* Krijg de tyfus en de griept.<sup>108</sup>  
           get the typhoid and the flu

We propose to ascribe the ungrammaticality of (66)c to this phenomenon. One cannot coordinate listed and non-listed items, even though this should be possible structurally.

In sum, we propose that differences between productive ANN compounds and lexicalized phrases can be derived from the fact that productive ANN compounds are interpreted compositionally and the second type is stored, even though they are built by means of the same structure. Both types contain an AP-layer.

### 5.3.5 Conclusion

In this section we have proposed that the non-head of lexicalized compounds contains nothing but bare roots. As such, we can derive the fact that they disallow adjectival features, such as degree modifiers, comparatives and superlatives and inflection. The fact that lexicalized compounds are obligatorily listed follows from the absence of a functional structure. As LF depends on functional structure to interpret the predicate expressed by a root compositionally, the conceptual module necessarily depends on readily available stored information.

We proposed that the left-hand part of productive ANN compounds and lexicalized phrases is a partial NP containing an AP-layer. As such, it follows immediately that these types may contain adjectival features, such as inflection and degree modification and that they are assigned DP stress. We have argued that restrictions on lexicalized phrases follows from the lexical property of being listed, rather than from structural considerations.

More generally, we have observed that being lexicalized is not associated with one particular structure. Being lexicalized is a property which cross cuts structural properties. It therefore does not make sense to postulate a separate module in which lexicalized compounds are derived (pace Ackema and Neeleman 2004 on ANN compounding). We have argued that all compounds are built in syntax, yet the level of merge may differ. Some compounds contain roots, whereas others contain phrases.

## 5.4. A previous account of type 1 ANN compounds

The data we have called ANN compounds of type 1 have been discussed before by Ackema and Neeleman (2004). They use these data to substantiate the claim that a morphological submodule blocks a syntactic derivation for listed items. As such, their proposal contrasts with our view according to which all ANN-compounds are derived in a single module, viz. syntax, and according to which there is no inherent relation between specific structures and listedness. As both the data and the claim are relevant for the present paper we will discuss their proposal in detail and we will argue in favor of the present proposal.

### 5.4.1 Ackema and Neeleman's (2004) proposal

Ackema and Neeleman (2004) propose that syntax and morphology are two separate submodules of narrow syntax. They further sub

---

<sup>108</sup> The example is ungrammatical as an imprecation.

mit that a syntactic derivation blocks the morphological derivation of syntactic objects unless the meaning of the morphological structure is not identical to the syntactic one, i.e. when the morphological structure is listed:

- (71) 'Let  $a_1$  and  $a_2$  be syntactic representations headed by  $a$ .  $a_1$  blocks  $a_2$  iff
- (i) in  $a_1$  (a projection of)  $a$  is merged with (a projection of)  $b$  in syntax, while in  $a_2$  a projection of  $a$  is merged with (a projection of)  $b$  in morphology', and
  - (ii) the semantic relation between  $a$  and  $b$  is identical in  $a_1$  and  $a_2$ .'(A&N 2004:51)

The examples in (72) illustrate their claim. They show ANN compounds of type 1 with a listed meaning.

- (72) a. zoet-hout      b. speciaal-zaak      c. zwart-boek      d. bruin-vis  
 sweet-wood      special-shop      black-book      brown-fish  
 'licorice'      'specialist shop'      'black book'      'porpoise'

Ackema and Neeleman (2004:62) point out that these examples do not have the same meaning as the corresponding syntactic phrases. The phrases in (73) contain the same nouns and adjectives as in (72). The AP occurs prenominally.

- (73) a. het zoete hout      b. de speciale zaak      c. het zwarte boek  
 the sweet.INFL wood      the special.INFL shop      the black.INFL book  
 'the sweet wood'      'the special shop'      'the black book'  
 d. de bruine vis  
 the brown.INFL fish  
 'the brown fish'

They conclude that listing blocks the syntactic derivation of these compounds. They further argue that listing requires non-compositional semantics (A&N 2004:83). They claim that, as a consequence, AN-compounds with a transparent meaning do not occur independently (see (74)a). They are blocked by their syntactic counterparts (see (74)b).

- (74) a. \* Zij heeft mooie blauw-oog-en.      (A&N 2004:63, ex.26)  
 She has beautiful.INFL blue-eye-s  
 b.    Zij heeft mooie blauwe ogen.      (A&N 2004:63, ex.26)  
 She has beautiful.INFL blue eyes  
 'She has beautiful blue eyes.'

Ackema and Neeleman (*ibidem*) note that there is an exception to the rule. When the structure will be embedded in a morphological structure, it will be derived by morphology, even if its meaning is transparent. For example, the affix *-ig* selects a morphological object. Due to this selectional requirement, the morphological structure *blauwoog* 'blue-eye' in (75) is not blocked by syntax.

- (75) Zij is blauw-oog-ig.

(A&N 2004:63, ex.25)

She is blue-eye-ADJECTIVAL.AFFIX  
 'She is blue-eyed.'

Similarly, it is possible to embed a transparent AN compound in another compound, as in (76) (the examples are taken from A&N 2004:83-84, ex.62). Ackema and Neeleman (2004:83) propose that in this case the morphological derivation of the AN structure is licensed because the complete compound in which it is contained is listed.

- |   |   |  |
|---|---|--|
| (76) a. kort-hoorn-koe<br>short-horn-cow<br>'breed of cow with short horns' | b. lang-poot-mug<br>long-leg-mosquito<br>'daddy longlegs' | c. lang-gat-boormachine<br>long-hole-drill<br>'drill for long holes' |
|---|---|--|

In sum, Ackema and Neeleman propose that syntax and morphology are two separate submodules of narrow syntax. They argue that syntactic merger blocks morphological merger. In other words, DP-formation blocks AN-compounds. However, they observe three exceptions. A compound may be derived when it merges with an affix that requires a morphological object or when it is the left hand part of another compound. Finally, a compound may be derived when it has a different meaning than the corresponding DP.

#### 5.4.2 Problems with the proposal

In this section we discuss some problems for Ackema's and Neeleman's (2004) claim that DPs block AN compounds with three exceptions. We discuss each exception in turn below.

Firstly, Ackema and Neeleman (2004) postulate that syntax blocks morphology except when the [AN] structure merges with an affix that requires a morphological object. This is a theory-internal argument. The statement that affixes (such as *-ig*) select for a morphological object is an assumption which depends on the view that syntax and morphology are separate (sub)modules. If one assumes but one module (i.e. syntax) it simply no longer follows. As such, this observation does not force us to assume two separate modules.

Secondly, they propose that syntax blocks morphology except when the [AN] structure is the left hand part of another compound, as in (77) (A&N 2004:83-84, ex.62).

- |  |   |  |
|--|---|--|
| (77) a. kort-hoorn-koe<br>short-horn-cow<br>'short horn' | b. lang-poot-mug<br>long-leg-mosquito<br>'daddy longlegs' | c. lang-gat-boormachine<br>long-hole-drill<br>'drill for long holes' |
|--|---|--|

Ackema and Neeleman (2004) explicitly state that the listedness of the compound as a whole licenses the morphological derivation of the transparent left-hand part. There are two problems with this particular proposal. The first problem is the fact that the right hand part of these compound may vary, as is illustrated in (78)-((80). The left-hand part is identical in the *a*, *b*, and *c* examples, but the right-hand part differs.

- |   |   |  |
|---|---|--|
| (78) a. kort-hoorn-rund<br>short-horn-bovine<br>'short horn bovine' | b. kort-hoorn-vee<br>short-horn-cattle<br>'short horn cattle' | c. kort-hoorn-veeras<br>short-horn-cattle.type<br>'short horn cattle type' |
|---|---|--|

- (79) a. lang-gat-bit  
long-hole-bit  
'bit for long holes'
- b. lang-gat-frees  
long-hole-milling.cutter  
'milling cutter for long holes'
- (80) a. vol-bloed-fokkerij  
full-blood-breeding.farm  
'breeding farm for thoroughbred animals'
- b. vol-bloed-paard  
full-blood-horse  
'thoroughbred horse'
- c. vol-bloed-stier  
full-blood-bull  
'thoroughbred bull'

One can even use these left hand part ANs to form new compounds productively, as shown in (81).

- (81) a. vol-bloed-veulen  
full-blood-foal  
'thoroughbred foal'
- b. vol-bloed-merrie  
full-blood-mare  
'thoroughbred mare'
- c. vol-bloed-hengst  
full-blood-stallion  
'thoroughbred stallion'
- d. vol-bloed-ruin  
full-blood-gelding  
'thoroughbred gelding'
- e. vol-bloed-ezel  
full-blood-donkey  
'thoroughbred donkey (ironic)'

The claim that the complete compound is listed is therefore false.

The second problem is the fact that the so-called left hand part does not need to be embedded at all. It is easy to find exocentric alternations on Google, as in (82) and (84). The examples in ((84)) are the exocentric versions of the compounds in (83).

- (82) a. kort-hoorn  
short-horn  
'short horn'
- b. lang-poot  
long-leg  
'daddy longlegs'
- c. lang-gat  
long-hole  
'drill for long holes'
- (83) a. vol-bloed-paard  
full-blood-horse  
'thoroughbred horse'
- b. lang-oor-konijn  
long-ear-rabbit  
'long-eared rabbit'
- c. breed-beeld-televisie  
wide-screen-television  
'wide screen TV'
- (84) a. vol-bloed  
full-blood  
'thoroughbred horse'
- b. lang-oor  
long-ear  
'long-eared rabbit'
- c. breed-beeld  
wide-screen  
'wide screen TV'

One could of course argue that the exocentric compounds are elliptical version of the endocentric ones. However, they select a different determiner, as can be seen in (85).

- (85) a. **het** vol-bloed<sub>NEUTER</sub>-paard<sub>NEUTER</sub>  
          the full-blood-horse  
          ‘the thoroughbred horse’

b. **het** lang-oor<sub>NEUTER</sub>-konijn<sub>NEUTER</sub>  
          the long-ear-rabbit  
          ‘the long-eared rabbit’

c. **de**<sub>COMMON</sub> breed-beeld<sub>NEUTER</sub>-televisie<sub>COMMON</sub>  
          the wide-screen-television  
          ‘the wide screen TV’

(86) a. **de**<sub>COMMON</sub> vol-bloed<sub>NEUTER</sub>  
          the full-blood  
          ‘the thoroughbred horse’

b. **de**<sub>COMMON</sub> lang-oor<sub>NEUTER</sub>  
          the-long-ear  
          ‘the long-eared rabbit’

c. **het**<sub>NEUTER</sub> breed-beeld<sub>NEUTER</sub>  
          the wide-screen  
          ‘the wide screen TV’

The gender of the determiner in the tripartite compounds in (85) is determined by the gender of the compound's head. The determiner simply agrees with the compound's rightmost part. In the bipartite compounds the issue is slightly more complicated. The determiner of a compound referring to an inanimate concept will be determined by the head of the compound as well. The compound referring to animate concepts, in contrast, shows agreement *ad sensum*. The determiner takes common gender as this gender is associated with animacy.

Technically, it is not clear where the gender assigned ad sensum comes from. One might of course argue that agreement ad sensum is derived through a null affix. Ackema & Neeleman (*ibidem*) could argue that this null affix is an affix that requires a morphological object to merge with. However, if one does so, one arrives at the same theory-internal argument as presented above. Affixes may only require morphological objects under the assumption that morphological objects exist. Under the proposal that there is but one generating module, viz. syntax, they will simply merge with a syntactic structure.

Thirdly, Ackema and Neeleman (2004) argue that syntax blocks morphology except when the [AN] structure is not transparent. Meaning is a theory-independent criterion. Yet, empirically the argument seems to be flawed. Even though [AN] compounds are always listed, their meaning may be transparent (see section 1).

- (87) a. diep-zee      b. zoet-stof      c. laag-land      d. hard-glas  
       deep-sea      sweet-matter      low-land      hard-glass  
       ‘deep sea’      ‘sweetener’      ‘lowland’      ‘hard glass’

They are further not necessarily blocked by a corresponding DP, as shown in ((88) and ((89)).

- (88) a. hoog-conjunctuur  
high-economic.climate  
'boom'  
b. een hoge conjunctuur  
a high economic.climate  
'a boom'

- (89) a. groot-vee  
large-cattle  
'cattle'
- b. het grote vee  
the large cattle  
'cattle'

In contrast, DPs may be idiomatic, as in (90).

- (90) a. een olijke Frans  
a jolly Frans  
'a jolly chap'
- b. een bittere pil  
a bitter pill  
'a painful experience'
- c. een harde tante  
a tough aunt  
'a tough cookie'
- d. een lange arm  
a long arm  
'a network connection'

Idiomaticity is thus not tied to compounds or DPs specifically. Whether a listed combination of an A and an N will be realized as a DP or as a compound is rather part of the listed information. This conclusion corresponds with what we have observed for type 2 and type 3 compounds in the present article. We have shown that even though they differ in listedness, they are identical from a structural point of view.

### 5.4.3 Theoretical consequences

In the previous section we have seen that Ackema & Neeleman's (2004) theory-independent criterion to make sure morphological merger can obtain, i.e. idiomaticity, is empirically flawed. If idiomaticity is not the criterion that regulates blocking between the modules, then how could these modules be ordered?

It is clear that free choice is too strong, as can be deduced from (91).

- (91) a. \* fris-wind  
fresh-wind
- b. # het zoete hout  
the sweet.INFL wood

One might propose yet another principle. However, we failed to find a single non-structural property that is unique to either compounds or DPs. As an alternative, we propose that syntax always wins. In other words, there is just one module. If a listed combination is expressed as a compound or a DP it is then simply part of the listed information.

### 5.5 Conclusion

In this article we have distinguished between three types of ANN compounds in Dutch, which we referred to as lexicalized compounds, lexicalized phrases and productive ANN compounds. We have shown that the structural properties of these types, which we derived from the level of merge, cross cut the lexical property of being lexicalized. Given that structural properties do not show a one-to-one mapping with lexical properties, it is undesirable to postulate an association between listedness and structure. Furthermore, we proposed that all AN(N) compounds are merged in syntax.

More specifically, we have shown that lexicalized compounds are fixed, listed combinations which do not allow any adjectival material, such as degree modification or adjectival inflection. They do not take DP stress either. We have argued that these properties result from the fact that they do not contain an adjective in a technical sense. What is recognized as an adjective is structurally but a bare, categoriless root. We have further derived the property of being lexicalized from this structure. Given that the root lacks adjectival functional projections, it cannot be interpreted as an adjective and LF fails to compute a compositional meaning. The structure depends on the availability of a stored denotation.

We have analyzed both compounds including lexicalized phrases and productive ANN compounds as compounds of which the non-head is a partial NP hosting an AP-layer. This structure captures the fact that the adjective in these compounds contain adjectival properties, such as inflection or degree modification and that the AN phrase receives DP stress. We have emphasized on the fact that restrictions on lexicalized phrases do not stem from structural distinctions, but from the mere fact that they are lexicalized. In sum, we have argued that conclusions on structure building in the domain of word-formation should not be based on lexical properties, such as being lexicalized.

## 6 Spectacular Splitters

### Posture verbs, motion verbs and copulas in Dutch and other West-Germanic varieties

#### Abstract

It has been observed in typological literature that posture verbs often grammaticalise into copulas (Newman 2002). The present article presents an account for this phenomenon by discussing the phenomenon in Dutch and other continental West-Germanic varieties, viz. High German, Low German and Frisian. More specifically, I contrast malleability with grammaticalisation in the empirical domain of posture verbs and motion verbs. These verbs can all be used as lexical verbs. However, they also show bleached uses, which look like copulas. For some verbs, this similarity with copulas merely stems from the fact that they can occur in non-agentive, stative contexts, which contrasts with the agentive activity reading in their lexical use. I conclude that these aspectual features thus do not inherently belong to the verb, but show malleability. Yet, other verbs indeed have grammaticalised into copulas.

#### 6.1 Introduction

An important debate on the nature of the syntax-lexicon interface is the question whether functional meaning is inherent to a lexeme and thus projected by that lexeme or whether, alternatively, functional meaning is solely derived from functional projections, i.e. from the syntax. Proposals to reduce syntactico-semantic meaning to functional projections are the Exo-Skeletal Model by Borer (2005a, 2005b, 2013a) and many proposals within Distributed Morphology (Halle and Marantz 1993, 1994). An important empirical test to determine which semantic-syntactic properties are inherently tied to a lexeme is malleability. For example, Borer (2005a) demonstrated that so-called count nouns are fully acceptable in mass noun structures. She concludes that the feature ‘count’ is not inherently tied to the lexeme, neither semantically or syntactically. The count structure is rather derived from the nominal syntactic projections. This approach has proven to be successful within the nominal domain as the noun often shows flexible behaviour.

For verbs, however, the situation is much more complex. Thanks to the monumental work by Levin and Rappaport Hovav on verbs (e.g. Levin 1993, Levin & Rappaport Hovav 1995, 2005, Rappaport Hovav & Levin 1998), the field is well aware of the existence of verb classes. As such, even many proposals within Distributed Morphology and other approaches that syntactically decompose verbal meaning into functional layers will accept or even emphasise on the fact that a verbal lexeme belongs to a class. A structure will only be acceptable if the verb’s functional projections match its lexical properties (e.g. Folli & Harley 2005, Wood 2015, Wood & Marantz 2016). The question which semantic properties are inherently tied to the verb and which are derived from functional projections is therefore still a central topic in the morphosyntactic debate and new observations and relevant empirical domains are continuously brought to the table.

The present article aims to contribute to this debate, by looking at how malleability contrasts with grammaticalisation in the empirical domain of posture verbs and motion verbs. The following verbs will be studied:

1) *posture verbs*

to sit  
to stand  
to lie  
to hang

2) *motion verbs*

to come  
to fall

All of them can be used as lexical verbs. However, they also show bleached uses, as shown in the following dutch examples:

3) A: **Staat** de ballerina op het podium? (lexical reading)  
stands the ballerina on the stage  
'Is the ballerina standing on the stage?'

B: Nee, ze **ligt** op het podium.  
no, she lies on the stage  
'No, she is lying down on the stage.'

A: **Staat** het danseresje van Degas in de VS? (bleached reading)  
stands the little.dancing.girl of Degas in the US  
'Is Degas' ballerina in the US?'

B: # Nee, ze **ligt** in de VS.  
no, she lies in the US  
'No, she is lying down in the US.'

The lexical reading refers to a bodily position and, as such, the verb can be contrasted with another verb that refers to a bodily position. The bleached reading vaguely refers to the bodily position: we know that statues tend to have an upward position and, hence, the use of the verb *staan* 'to stand' is motivated. However, the meaning of the verb merely refers to a presence at a location, rather than referring explicitly to the bodily position. A contrast with another verb is odd.

Such bleached uses are attested in continental West-Germanic languages. Included in the present study are Dutch, High German (i.e. Standard German), Low German and Frisian, with a focus on Dutch, as Dutch seems to be the front-runner in this process. In what follows, I will use the English infinitive as an abstract name for the various variations of a verb. For example, when I speak about *to hang* I refer to the set of verbs that are the translations of *to hang* in these languages. Whether the verbs indeed show bleached uses depends both on the verb and the language, as summarised in the following table:

has a locational meaning in...	Dutch	Frisian	Low German	High German
to sit	✓	✓	✓	✗
to stand	✓	✓	✓	✓
to lie	✓	✓	✓	✓
to hang	✓	✓	✓	✓
to come	✓	✓	✓	✓
to fall	✓	✗	✗	✗

I argue that these bleached uses either result from malleability or grammaticalisation in continental West-Germanic languages. More specifically, I will argue that in Dutch *zitten* 'to sit', *komen* 'to come' and *vallen* 'to fall' are truly grammaticalised into copulas. An example of a bleached reading *because of malleability* is given in (4). In such uses, the verbs occur in non-agentive, stative contexts. In such contexts, the verb loses its agentive meaning and its *Aktionsart* as an activity, which it does have in its core lexical use, such as in *The acrobat is standing on his head*. I will conclude that these properties thus exist independently from the root, as the root can occur in both contexts. They thus show malleability:

- 4) De Eiffeltoren staat in Frankrijk.  
 the Eiffel.tower stands in France  
 'The Eiffel Tower is in France.'

Copulas result from true grammaticalisation: the language's vocabulary gains an extra functional vocabulary item, i.e. a copula, alongside the original root, which may still occur as a lexical verb. An example of a *copular use* is illustrated in (5):

- 5) Anna zit in Frankrijk.  
 Anna sits in France  
 'Anna is in France.'

To get a feeling of where the present contribution is leading to, note that one can immediately notice a significant difference between the bleached reading through malleability and the one through grammaticalisation into a copula: the copula allows for an animate subject (5), the non-agentive stative reading does not (4).

I will conclude that the verbs should be grouped as follows:

- 6) *stative use of the lexical verb, i.e. the result of malleability*  
 to stand (posture verb)  
 to lie (posture verb)  
 to hang (posture verb)

7) *true copulas, i.e. the result of grammaticalisation*

- to sit (posture verb)
- to come (motion verb)
- to fall (motion verb)

Malleability is available in all languages under discussion, grammaticalisation into copulas is a language-specific phenomenon. Given that the malleability of the verbs *to stand*, *to lie* and *to hang* is available in all languages, it immediately follows that we attest the bleached uses in all languages. *To come* happened to grammaticalise into a copula in all varieties, *to sit* did so in most varieties, but not in High German and *to fall* only did so in Dutch. This summary will account for the observations in table 1.

It is interesting to investigate which semantic properties are retained and which are lost when the lexical verb is used with non-canonical functional projections (in the case of malleability) and when the verb loses its lexical meaning altogether (in the case of grammaticalisation). The phenomenon of *malleability* occurs for these verbs as they can be used in a non-agentive, stative way. We will see that the resulting syntactic structure is reminiscent aspectually of the syntactic contexts of copulas. However, the verbs are not copulas in the strict sense of the word.

I start by using Dutch examples to establish the main distinctions within the discussion , even though I promised an article on continental West-Germanic languages. It will soon become clear to the reader why I do so: the discussion is empirically rather complex and restricting the discussion to a single language is more than convoluted enough. The discussion on Dutch is representative of the main distinctions in these continental West-Germanic languages. After having established the main distinctions in the verbal domain under discussion, I will turn to a comparative perspective again.

Before we proceed, I would like to mention what the present article is not about. Firstly, it is well known amongst Dutch linguists that posture verbs can also serve as aspectual auxiliaries (see Broekhuis 2015: chapter 6; Pots 2020). This grammaticalisation process seems to be a typologically well attested property of posture verbs as well (see Newman 2002). Needless to say, it is an important and relevant observation, but it is not on the agenda for today. Neither is the present article about when ones uses which posture verb in Dutch. The question in which context, for example, *zitten* ‘to sit’ is to be preferred over *staan* ‘to stand’ in Dutch is highly intriguing and complex, but this article does not aim to address the question directly. I refer the reader to work by Van den Toorn (1975) and Lemmens (2002).

The article is structured as follows. The following section sets the stage by briefly discussing typological observations on copula inventories, which shows that it is possible for a language to have an array of copulas. A next section discusses the various *readings* of the Dutch posture verbs and motion verbs. Admittedly, it will become clear that these various readings are less relevant for the syntactic distinctions that will later be proposed in the article. However, they do help to disentangle the somewhat confusing empirical domain of Dutch posture and motion verbs. Section 4 then discusses which functional heads can be associated with which readings. We will see that the verbs can occur with various sets of functional heads and I will conclude that these features thus do not inherently belong to the root. Furthermore, these functional distinctions intersect the reading-based classification. We will see that three verbs in particular seem to be behave on a par with the copula *zijn* ‘to be’. These verbs are *zitten*, *komen* and *vallen*. This raises the question whether they are perhaps truly copulas. Sections 5 shows that, indeed, they are copulas and that they, as such, do

differ from the other posture verbs and that they tend to be copulas not only in Dutch, but in West-Germanic varieties more generally. The article then shifts its focus fully towards the semantics and the syntax of these copulas in section 6. Section 7 discusses some semantic properties of the copulas and section 8 expands on their syntax. Section 9 concludes.

## 6.2 Sharers and splitters

When it comes to the distribution of copulas cross-linguistically, the typological literature distinguishes between so-called ‘sharers’ and ‘splitters’(Stassen 2009:266). Sharers are languages for which the copula for locational predication and nominal predication is the same. In contrast, splitters are languages for which the copula for locational predication differs from the one for nominal predication. An example of a sharer is given in (8) (example taken from Stassen (2009:266), an example of a splitter is shown in (9) (example taken from Zaguna (2012:303)).

- 8) a. Maria hayko-v-č yu-m. (Yavapai)  
           Maria Anglo-DEM-SUBJECT be-ASPECT  
           ‘Maria is an Anglo.’
- b. Cnapuk-č miyul yu-m.  
       ant-SUBJECT sugar-in be-ASPECT  
       ‘The ant is in the sugar.’
- 9) a. El siete es un número impar. (Spanish)  
       the seven is a number odd  
       ‘Seven is an odd number.’
- b. El coche está en el garaje.  
       the car is in the garage  
       ‘The car is in the garage.’

The example in (8) shows that Yavapai uses the same copula for both nominal and locational predication, (9) illustrates that Spanish has *ser* with nominal predication, whereas it has *estar* for locational predication.

Now consider Dutch predicational constructions. The default copula for nominal predication in Dutch is *zijn* ‘to be’, as shown in (10).<sup>109</sup>

- 10) Jan is een grote man.  
     John is a tall man  
     ‘John is a tall man.’

However, in the case of constructions that at least *look like* locational predication, a plethora of verbs can occur. *Zijn* ‘to be’ itself is available, as in (11), alongside the posture verbs *zitten* ‘to sit’, *staan* ‘to

---

<sup>109</sup> Just like other languages, Dutch also has a non-stative copula *worden* ‘to become’.

stand', *liggen* 'to lie' and *hangen* 'to hang', see (12), and two motion verbs, i.e. *vallen* 'to fall' and *komen* 'to come', as shown in (13).

- 11) Jan is in Frankrijk.

John is in France

'John is in France.'

- 12) a. Jan zit in Frankrijk.

John sits in France

'John is in France.'

- b. De Eiffeltoren staat in Frankrijk.

the Eiffel.tower stands in France

'The Eiffel Tower is in France.'

- c. Mijn fiets staat achter de muur.

My bike stands behind the wall

'My bike is standing behind the wall.'

- d. Parijs ligt in Frankrijk.

Paris lies in France

'Paris is in France.'

- e. Het wrak/de boot ligt bij Larnaca.

the wreck/the boat lies near Larnaca

'The wreck/the boat lies near Larnaca.'

- f. De lamp hangt aan het plafond.

the lamp hang on the ceiling

'The lamp hangs on the ceiling.'

- 13) a. Die oorlog viel in de 20<sup>ste</sup> eeuw.

that war fell in the 20<sup>th</sup> century

'That war is situated in the 20<sup>th</sup> century.'

- b. Ze komt uit Frankrijk.

she comes from France

'She is from France.'

In all of the above examples the verb could be replaced by the default copula *zijn* 'to be' without meaning loss. The question can be raised whether this means that Dutch is a splitting language and whether all of the above verbs then grammaticalised into copulas. In the following sections I will argue that *zitten*, *vallen* and *komen* are indeed grammaticalised into copulas, just as are some of

their equivalents in the other continental West-Germanic languages.<sup>110</sup> The other verbs are not true copulas, but involve aspectually shifted and semantically bleached uses of the lexical verbs in all continental West-Germanic languages under discussion.

In what follows, I will refer to examples such as the ones above as *locational readings*. Strictly speaking, this is not entirely accurate for *vallen* ‘to fall’, which has a temporal reading, and *komen* ‘to come’, which has a reading that is more directional – it expresses an origin- than locational, but I will ignore these details when they are not directly under discussion for ease of exposition. When the verbs are used in their original sense, as in *the acrobat is standing on his head*, I will call this use the *lexical reading*.

### 6.3 Distinguishing between lexical readings and locational readings

#### 6.3.1 Introduction

The verbs under discussion can all still be used as lexical verbs of posture and verbs of motion, alongside the locational readings. For motion verbs, it is very easy to set apart the two readings. Under the lexical reading, they have the familiar interpretations of ‘falling down’ for *vallen* ‘to fall’ and of ‘arriving/attending’ for *komen* ‘to come’:

- 14) Jan viel uit de boom.  
John fell from the tree  
'Jan fell out of the tree.'
- 15) Komt je vrouw naar het feestje?  
comes your wife to the party  
'Is your wife coming to the party?'

Yet, under the locational readings, they get a very specific interpretation: *vallen* ‘to fall’ receives a temporal reading and *komen* ‘to come’ gets an origin reading:

- 16) Die oorlog viel in de twintigste eeuw.  
that war fell in the twentieth century  
'That war took place in the twentieth century.'
- 17) Komt je vrouw uit Frankrijk?  
comes your wife from France  
Lexical reading: 'Is your wife arriving from France?'  
Locational (i.e. origin) reading: 'Is your wife from France, i.e. is she French?'

As such, it is not hard to set apart the lexical reading from the locational one for these verbs. Admittedly, the example in (17) is ambiguous, which does complicate matters a bit, but the

---

<sup>110</sup> Surprisingly enough, I was not able to find any source in which this observation has been mentioned previously. Hans Broekhuis (p.c.), the author of the extensive work *The Syntax of Dutch*, confirmed he is not aware of any previous literature either.

difference between the two different readings is semantically salient enough to set them apart easily.

However, distinguishing between the lexical reading and the locational one is much harder for the posture verbs. Consider the example below, which is perfectly ambiguous:

- 18)a. Anna zit in de tuin.  
Anna sits in the garden.  
Lexical reading: 'Anna is sitting in the garden.'  
Locational reading: 'Anna is in the garden.'
- b. Anna zit in haar bed.  
Anna sits in her bed.  
Lexical reading: 'Anna is sitting in her bed.'  
Locational reading: 'Anna is in her bed.'

Given the ambiguity and the closeness in meaning, it may seem puzzling sometimes to set apart the two different readings in certain sentences. The following sections provide tools to approach this problem.

### 6.3.2 Distinction in grounds

There is a way to manipulate which reading will prevail. The size of the ground namely influences which reading will be triggered: the more spacious the ground, the more prominent the locational reading will be and vice versa. Consider the following examples:

- 16)a. Anna zit op de kruk.  
Anna sits on the stool.  
Lexical reading: 'Anna is sitting on the stool.'  
\* Locational reading: 'Anna is on the stool.' (Imagine she is hanging over the stool.)
- b. Anna zit op de tak.  
Anna sits on the branch.  
Lexical reading: 'Anna is sitting on the branch.'  
\*? Locational reading: 'Anna is on the branch.' (Imagine she is hanging over a branch.)
- 17)a. Anna zit in de tuin.  
Anna sits in the garden.  
Lexical reading: 'Anna is sitting in the garden.'  
Locational reading: 'Anna is in the garden.' (Imagine she is walking around.)
- b. Anna zit in haar bed.  
Anna sits in her bed.  
Lexical reading: 'Anna is sitting in her bed.'  
Locational reading: 'Anna is in her bed.' (Imagine she is lying down.)

- 18) Anna zit in Frankrijk.

Anna sits in France.

\*? Lexical reading: 'Anna is sitting in France (and not standing).'

Locational reading: 'Anna is in France.'

In example (16)a the stool is a small ground. It is probably smaller than a grown human body. For such an example, the lexical reading is so prominent, the locational reading is excluded. Note that this effect cannot be ascribed solely to the fact that a stool is semantically closely connected to sitting. The branch in the b-example has a similar effect.

In (17), the grounds, i.e. the garden and the bed, are larger than a human body. However, they are still relatively small in the sense that it would be possible to observe an entire garden or a complete bed in one gaze. Such grounds trigger highly ambiguous readings and the interpretation of the sentence will depend on the context.

In (18), in contrast, the ground, viz. France, is so spacious that one can only grasp the entire space conceptually. It is impossible to observe France in its entirety within a single human gaze. Such spaces strongly trigger the locational reading. Admittedly, I do not think the lexical reading is fully excluded. If this world had sitting competitions and if we know that Anna travels the world to compete as a professional athlete in this discipline, I guess that (18) might trigger the lexical interpretation. However, it should be clear that this reading is highly marked and generally not available. In sum, the more spacious the ground, the more salient the locational reading will be and vice versa.

The observation does not only hold for *zitten* 'to sit', but also for the other postures verbs. Consider the following two examples, which show the verb *liggen* 'to lie':

- 19)a. Het boek ligt op die kruk.

the book lies on that stool

Lexical reading: 'The book is lying down on that stool.'

- b. Het boek ligt in mijn kast.

The book lies in my cabinet

Lexical reading: 'The book is lying down in my cabinet.'

Locational reading: 'The book is in my cabinet.'

- c. Het boek ligt in Frankrijk.

the book lies in France

Locational reading: 'The book is in France.'

The smaller the ground, the more salient is the implication that the posture expressed by the verb is to be taken literally. For the *stool*-example, I would be really surprised to see that the book is

actually standing, for the *cabinet*-example it would not have been my first assumption, but I would accept it and for the *France*-example I have no expectations whatsoever on the posture of the book.

### 6.3.3 Distinction in meaning

The verbs mean something else when used lexically and locationally. The lexical reading of the example in (20) quite literally implies the action denoted by the verb. Under the lexical reading, Anna is literally sitting in the garden. The locational reading has a different semantics. Under this reading it is merely implied that Anna is in the garden.

- 20) Anna zit in de tuin.  
Anna sits in the garden.  
Lexical reading: 'Anna is sitting in the garden.'  
Locational reading: 'Anna is in the garden.'

The ambiguity can be detected by rephrasing the sentence. When used locationally, the verb can be replaced by the copula *zijn* 'to be', when used lexically, the verb's paraphrase may involve words associated with the semantic field of the action in its most literal sense. Consider, for example, the following dialogue. A valid answers under the lexical reading of *zitten* 'to sit' involves elements such as a garden chair:

- 21) Zit Anna in de tuin?  
Sits Anna in the garden?

*Licit reply under the copula reading:*

Ja, ze **is** in de tuin. Ze zei dat ze nog werk had aan de tomatenplanten.  
'Yes, she is in the garden. She said she had work to do on the tomato plants.'

*Licit reply under the lexical reading:*

Ja, ze **zit** neer. Ik kan zien dat ze de nieuwe tuinstoel gebruikt.  
'Yes, she is sitting down. I can see she is using the new garden chair.'

Furthermore, under a lexical reading, the verbs of posture can often be contrasted by means of another lexical verb of posture, as shown in (22):

- 22) A: Ligt de hond in zijn mand?  
lies the dog in his basket  
'Is the dog lying in his dog bed?'  
  
B: Nee, hij **zit** in zijn mand.  
no, he sits in his basket  
'No, he is sitting in his dog bed.'

In contrast, under the locational reading, no specific posture or motion is implied. For example, the sentence in (23) does not imply a specific bodily posture. It may well be the case, for example, that Anna is walking around. The only fact that is implied is her location, not her bodily posture:

- 23) Anna zit in Parijs.  
Anna sits in Paris  
'Anna is in Paris.' (She may be walking around.)

As a result, the following conversation is odd:

- 24) A: Zit Anna in Parijs?  
           sits Anna in Paris  
           'Is Anna in Paris?'

B: # Nee, ze **ligt** in Parijs  
No, she lies in Paris

What is licit, though, is the following response:

- 25) B: Ja, ze loopt lekker te winkelen in de Marais.  
yes, she walks pleasantly to shop in the Marais  
'Yes, she is shopping enjoyably in the Marais.'

In short, under the lexical meaning the bodily posture expressed by the verb is to be taken literally. No bodily posture is implied by the verb under the locational reading.

### 6.3.4 Causativity

First note that for each posture verb there is a corresponding causative verb in Dutch:

- |     |               |            |                |                                |
|-----|---------------|------------|----------------|--------------------------------|
| 26) | <i>zitten</i> | 'to sit'   | <i>zetten</i>  | 'to put'                       |
|     | <i>liggen</i> | 'to lie'   | <i>leggen</i>  | 'to lay'                       |
|     | <i>staan</i>  | 'to stand' | <i>stellen</i> | 'to position'                  |
|     | <i>hangen</i> | 'to hang'  | <i>hangen</i>  | 'to hang' <sub>CAUSATIVE</sub> |

Each lexical reading can be understood as resulting from the corresponding causative version. For animate, volitional subjects, the understood *causer* can be the subject itself or someone else, as shown in (27). For non-animate subjects, it is understood that the posture is caused by an unmentioned, volitional agent, as in (28):

- 27) De hond ligt in zijn mand. (posture verb)  
the dog lies in his basket  
'The dog is lying in his dog bed.'

Hij heeft zich daar gelegd. / Ik heb hem daar gelegd. (causative verb)  
he has REFL there laid/ I have him there laid  
'He has laid himself down there./ I put him there.'

- 28) Het boek ligt in mijn kast. Ik heb het daar gelegd.  
The book lies in my cabinet I have it there laid  
'The book is lying in my cabinet. I put it there.'

Note that the subject, i.e. the dog, in (27) is thus interpreted as the agent of the lying event, but not necessarily as the causer of the event. I will come back to this in section 4.2, in which I argue that the lexical reading contains Voice<sub>AGENT</sub>.

Locational readings do not have this causal implication. As such, they allow for readings in which no cause for the position of the subject can be given. For example, no one put Normandy in France, as in (29) or put caffeine in coffee, as in (30). These positions are simply given.

- 29) Normandië ligt in Frankrijk.  
 Normandy lies in France  
 'Normandy is in France.'

- 30) Er zit coffeïne in koffie.  
 There is caffeine in coffee  
 'There is caffeine in coffee.'

However, even when world knowledge allows one to think of a causer, locational readings resist causative alternants. The causative nature of the sentence in (32) triggers a lexical reading, which is incompatible with the spacious ground, resulting in an odd sentence.

- 31) De Mona Lisa hangt in Frankrijk.  
 The Mona Lisa hangs in France  
 'The Mona Lisa is in France.'

- 32) # Ze hebben de Mona Lisa in Frankrijk gehangen.  
 they have the Mona Lisa in France hung  
 'They've hung the Mona Lisa in France.'

In sum, lexical readings and locational readings of posture verbs can be distinguished by the fact that there is an understood causative event for lexical readings, which is not available for locational readings.

### 6.3.5 Choice of posture verb

Under lexical readings, the choice of the verb is not given, but merely reflects the posture the subject happens to have. For example, under the lexical reading the dog may be sitting, lying or standing in his dog bed and a book can be lying down or standing upright on my shelf. For locational readings, however, the verb is given. It is mostly determined by the shape of the subject (see Van den Toorn 1975, Lemmens 2002). Native speakers of Dutch know that towers and statues are standing, that a sea is 'lying' and that paintings are hanging, etc. Choosing the non-matching verb in a locational reading will trigger a lexical reading for smaller grounds, such as *the table* in (33). The exact posture is then implied: if the painting happens to be lying down on the table, the sentence will be seen as false. Alternatively, it will result in non-interpretability for larger grounds, such as France in (34) or the hearer accommodates his/her world view to interpret the sentence. For the example in which it is claimed that the Mona Lisa is standing in France, the hearer will either fail to interpret the

sentence or they will conclude that they are perhaps mistaken in thinking that the Mona Lisa is a painting rather than a statue:

- 33) De Mona Lisa staat op de tafel.  
the Mona Lisa stands on the table  
'The Mona Lisa is standing on the table.'

- 34) De Mona Lisa staat in Frankrijk.  
the Mona Lisa stands in France  
'The Mona Lisa is standing in France.'

In short, lexical readings allow the speaker to choose the posture verb, for locational readings the appropriate posture verb is given.

### 6.3.6 Overview

We now have established ways to distinguish between the lexical reading and the locational reading. Let us now check for each verb how they behave under the two readings. First consider the posture verbs. As expected, they can all be used lexically if the location is a small ground. This can be concluded from the fact that we can deny the bodily posture entailed by the verb:<sup>111</sup>

- 35) A: **Zit** de hond in zijn mand?  
sits the dog in his basket  
'Is the dog sitting in his dog bed?'

B: Nee, hij **ligt** in zijn mand.  
nee he lies in his basket  
'No, he is lying in his dog bed.'

- 36) A: **Staat** de vaas op tafel?  
stands the vase on table  
'Is the vase standing on the table?'

B: Nee, de vaas **ligt** op tafel.  
no the vase lies on table  
'No, the vase is lying on the table.'

- 37) A: **Ligt** het boek in je kast?  
lies the book in your cabinet  
'Is the book lying down in your cabinet?'

---

<sup>111</sup> The examples all allow for a locational reading as well, but that is not what I intend them to demonstrate, here.

B: Nee, het staat in mijn kast.  
no it stands in my cabinet  
'No, it is standing upright in my cabinet.'

- 38) A: **Hangt** het schilderij boven de eettafel?  
hangs the painting above the dining.table  
'Is the painting hanging above the dining table?'

B: Nee, het staat op de eettafel.  
no, it stands on the dining.table  
'No, it is standing on the dining table.'

The verbs can all be used locationally as well. In the examples below, this reading is triggered by the spacious grounds. Furthermore, in each example the verb is selected that is associated with the most prototypical posture of the subject involved. For example, a painting will hang prototypically, where a statue will stand upright. The selection of the most prototypical, default verb also induces the locational reading:

- 39) A: **Zit** je vrouw in Frankrijk?  
sits your wife in France  
'Is your wife in France?'

B: # Nee, ze ligt in Frankrijk.  
no, she lies in France  
'No, she is lying down in France.'

- 40) A: **Hangt** het schilderij in Frankrijk?  
hangs the painting in France  
'Is the painting in France?'

B: # Nee, het staat in Frankrijk.  
no, it stands in France  
'No, it is standing in France.'

- 41) A: **Ligt** het boek in Frankrijk?  
lies the book in France  
'Is the book in France?'

B: # Nee, het staat in Frankrijk.  
no it stands in France  
'No, it is standing upright in France.'

- 42) A: **Staat** het danseresje van Degas in de VS?  
stands the little.dancing.girl of Degas in the US  
'Is Degas' ballerina in the US?'

- B: # Nee, ze ligt in de VS.  
no, she lies in the US  
'No, she is lying down in the US.'

In sum, we can conclude that the posture verbs *zitten* 'to sit', *staan* 'to stand', *hangen* 'to hang' and *liggen* 'to lie' and the motion verbs *vallen* 'to fall' and *komen* 'to come' all occur under both lexical and locational readings. However, we will see that this is not ascribed to a single syntactic distinction. The following sections further disentangles the various occurrences of these verbs in the lexical and locational readings.

#### 6.4 The *Aktionsart* of posture verbs, motion verbs and copulas

##### 6.4.1 Introduction

The previous sections distinguished between lexical and locational readings. However, the inventory of these verbs is much more complex. Consider the following data. In the previous sections, I have said that whenever the bodily posture in the sentence can be denied or contrasted, we are observing a lexical reading. Such is the case for the following two examples:

- 43) A: Ligt de hond in zijn mand?  
lies the dog in his basket  
'Is the dog lying in his dog bed?'

- B: Nee, hij zit in zijn mand.  
nee he sit in his basket  
'No, he is sitting in his dog bed.'

- 44) A: Ligt het boek in je kast?  
lies the book in your cabinet  
'Is the book lying down in your cabinet?'

- B: Nee, het staat in mijne kast.  
no it stands in my cabinet  
'No, it is standing upright in my cabinet.'

However, the sentences seem to differ somehow. In (43), it seems like the dog is actually doing something (he is sitting), which is not true for the book in (44). To address this intuition, the following sections add extra distinctions to the classification. We will look at the syntactic properties of agentivity and unboundedness, i.e. durativity. We will see that this syntactic classification intersects the lexical-locational classification we arrived at above. For ease of exposition, I discuss posture verbs and motion verbs in separate sections. In later subsections, I also compare the posture verbs and the motion verbs with the copula *zijn* 'to be'. I will conclude this section with the conjecture that perhaps yet a third classification should be added to the picture, i.e. the one between copulas and non-copulas. Later sections will confirm this conjecture to be on the right track.

#### 6.4.2 Posture verbs in lexical readings as activities and states

Let us consider the contrast observed in the previous subsection: when a dog is lying or sitting, it feels like he is doing something (albeit something lazy), but when a book is lying down or standing upright, this is not felt as an activity, whereas both examples contain lexical readings. In the present section, I argue that this contrast is a syntactic contrast: the dog-example contains a syntactic activity, whereas the book-example is a syntactic state.

First note that the examples contrast in agentivity (see Alexiadou, Anagnostopoulou and Schäfer 2015: 20-21 for an overview and discussion of the agentivity tests used here). The agentive *dog*-example allows for adverbs modifying the intention of the agent, such as *deliberately*, it allows for subclauses of agentive purpose and by *itself*<sup>n2</sup> modification, whereas the *book*-example does not:

- 45) a. De hond ligt opzettelijk in zijn mand.  
the dog lies deliberately in his basket  
'The dog is lying deliberately in his dog bed.'
- b. De hond ligt in zijn mand om te slapen.  
the dog lies in his basket in.order to sleep  
'The dog is lying in his dog bed to sleep.'
- c. De hond lag uit zichzelf in zijn mand.  
the dog lies by himself in his basket  
'The dog went lying in his dog bed by himself.'
- 46) a. \* Het boek ligt opzettelijk in mijn kast.<sup>n3</sup>  
the book lies deliberately in my cabinet  
Intended: 'The book is lying deliberately in my cabinet.'
- b. \* Het boek ligt in mijn kast om zich te beschutten tegen stof.  
the book lies in my cabinet in.order REFL to protect against dust  
Intended: 'The book is lying in my cabinet to protect itself against dust.'
- c. \* Het boek lag uit zichzelf in de kast.  
the book lied by itself in the cabinet  
Intended: 'The book was lying by itself in the cabinet.'

Admittedly, agentivity tests are not always trustworthy for sentences which clearly contrast in animacy and volition. It is indeed hard to tell whether the agentivity tests show that the dog-example is agentive in the syntactic sense of the word, whereas the book-example is not or whether

---

<sup>n2</sup> In this test, *by itself* should be interpreted as 'by its own doing', not as 'alone'.

<sup>n3</sup> Not fully excluded is a reading in which someone put a book in my cabinet deliberately. However, what is fully excluded is that the subject of the standing event, i.e. the book, is doing this standing deliberately and that is what this example aims to show.

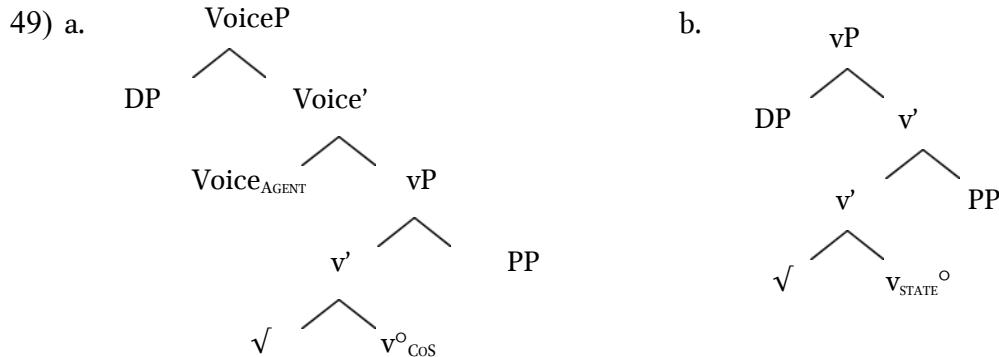
it simply indicates that the dog is a volitional subject, unlike the book.<sup>114</sup>

However, note that the agentivity distinction goes hand in hand with a distinction in durativity. The agentive examples do allow for a durational reading, whereas the non-agentive one does not:

- 47) De hond is in zijn mand aan het liggen.  
      the dog is in his basket on the lie.INFINITIVE  
      'The dog is lying in his dog bed.'
- 48) \* Het boek is in mijn kast aan het liggen.  
      the book is in my cabinet on the lie.INFINITIVE

The licitness of the durative reading of the *dog*-example indicates that the example contains an activity syntactically. Its illicitness for the *book*-example shows that the *book*-example contains a state syntactically. It is known that activities and states have very different subjects: the subject of activities are agents, whereas the subject of a state is merely the holder of the state (Kratzer 1996).

Alexiadou, Anagnostopoulou and Schäfer (2015: 48) propose that little v comes in two versions: one that introduces unbounded events for change-of-state (CoS) verbs and one that introduces stative events. I adopt this proposal and I assume that the agentive, durative lexical reading contains the unbounded little v, whereas the non-agentive, non-durative lexical reading contains the stative little v. Furthermore, I assume that the activity reading projects an agent syntactically, by means of a *Voice<sub>AGENT</sub>* head. The state does not contain such a projection. The tree in (49)a shows the structure of posture verbs as activities, the one in (49)b shows their structure as states. In these structures, the posture verb will realize the root that merges with little v.



The lexical reading thus actually hosts two different reading for the posture verbs, depending on the functional heads it merges with. I conclude that these heads are not intrinsically tied to the roots, given the observed malleability.

<sup>114</sup> Alexiadou, Anagnostopoulou & Schäfer (2015) point out that the *by itself*-test is not sensitive to animacy. This is certainly true when the test tests causation. It is in my opinion not always true if it tests for agentivity without causation (see below).

### 6.4.3 Locational readings of posture verbs are always states

For the locational reading of posture verbs, *staan* ‘to stand’, *liggen* ‘to lie’ and *hangen* ‘to hang’ differ as a group from *zitten* ‘to sit’. I will therefore limit this subsection to *staan* ‘to stand’, *liggen* ‘to lie’ and *hangen* ‘to hang’, postponing a discussion of *zitten* ‘to sit’ to the next subsection.

Consider the locational readings of *staan* ‘to stand’, *liggen* ‘to lie’ and *hangen* ‘to hang’. Unsurprisingly, these readings invariably fail all agentivity tests. I illustrate three tests for *staan* ‘to stand’ and I give one test for the other two verbs for reasons of brevity:

- 50) \* De Eiffeltoren staat opzettelijk in Parijs.<sup>115</sup>  
the Eiffel.tower stands deliberately in Paris
- 51) \* De Eiffeltoren staat in Parijs om toeristen te lokken.<sup>116</sup>  
the Eiffel.tower stands in Paris in.order tourists to attract
- 52) \* De Eiffeltoren staat uit zichzelf in Parijs.  
the Eiffel.tower stands by itself in Paris
- 53) \* Normandië ligt opzettelijk in Frankrijk.  
Normandy lies deliberately in France
- 54) \* De Mona Lisa hangt opzettelijk in Parijs.  
the Mona Lisa hangs deliberately in Paris

It is hard to tell what these tests show for these verbs as the subjects are invariably and obligatorily inanimate. As such, an agentive interpretation is hard to get. Alexiadou, Anagnostopoulou, Schäfer (2015) show that inanimate subject can nevertheless function as the causes of the event. The *by itself* test can respond to this notion, even when the subject is inanimate. Yet, we have seen in section 3.4 that the locational readings strictly exclude causative implications. Hence, it comes as no surprise that an internal causer is not available either. Taken together, the fact that these sentences resist an agentive reading might be a trivial observation. However, note that they also resist durative readings:

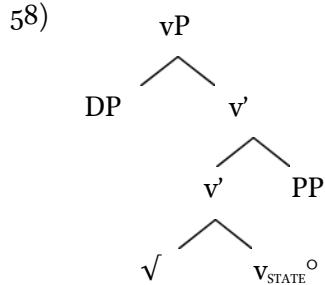
- 55) \* De Eiffeltoren is in Parijs aan het staan.  
the Eiffel.tower is in Paris on the stand.INFINITIVE
- 56) \* Normandië is in Frankrijk aan het liggen.  
Normandy is in Paris on the lie.INFINITIVE
- 57) \* De Mona Lisa is in Parijs aan het hangen.  
the Mona Lisa is in Paris on the hang.INFINITIVE

---

<sup>115</sup> Again, not fully excluded is a reading in which someone put the tower in Paris deliberately, but, again, that reading is not under discussion.

<sup>116</sup> Again, not fully excluded is a reading in which someone put the tower in Paris deliberately, but, again, that reading is not under discussion.

Together with the complete lack of agentivity, the lack of a durative reading shows that these readings are states as well. As such, the locational readings do not differ syntactically from those lexical readings which are states. I therefore propose that the locational reading of these verbs contain the stative little v and no Voice head:



The question then arises what triggers the difference between the lexical stative reading and the locational reading. In section 3.2 we have seen that the size of the ground can influence the salience of these readings. I conjecture it is therefore a pragmatic distinction, not a syntactic one.

#### 6.4.4 The locational reading of *zitten* ‘to sit’

When the subject is animate, *staan* ‘to stand’, *hangen* ‘to hang’ and *liggen* ‘to lie’ cannot get the locational reading, but only the lexical reading:

- 59) Jan staat/hangt/ligt in de tuin.  
 John stands/hangs/lies in the garden  
 ‘John is standing/hanging/lying in the garden.’ (lexical reading)  
 \* ‘John is in the garden.’ (locational reading)

Recall from section 3.2 that spacious grounds trigger a locational reading. If we now combine an animate subject and a posture verb other than *zitten* with a spacious ground the result is not ungrammaticality, but metonymy. The animate subject is reinterpreted as if it were inanimate:

- 60) Mijn man staat in Frankrijk.  
 my husband stands in France  
 Only locational reading: ‘My husband (in fact, the caravan under control of my husband) is standing in France.’

This example shows that the structure forces the subject to be interpreted as an inanimate entity.

Importantly, the verb *zitten* ‘to sit’ behaves differently in this respect. It can get the locational readings when the subject is animate:

- 61) Jan zit in Frankrijk.  
 John sits in France  
 ‘John is in France.’

Note, for good measure, that *zitten* ‘to sit’ *can* have animate subjects, but it does not *have* to have animate subjects, as illustrated in (62) below:

- 62) Er zit suiker in de koffie.  
there sits sugar in the coffee  
'There is sugar in the coffee.'

Consider further that the verb *zitten* in its locational reading clearly resists a durative reading:

- 63) \* Anna is in Parijs aan het zitten.  
Anna is in Paris on the sit.INFINITIVE  
Intended: 'Anna is in Paris for a while.'

This is what we expect: we have seen that locational readings are states. Locational readings with *zitten* ‘to sit’ do not seem to differ: they behave as states. I conclude it has the stative little v.

However, when it comes to agentivity, it passes all tests with flying colours if the agent is animate:

- 64) a. Anna zit opzettelijk in Parijs.  
Anna sits deliberately in Paris  
'Anna is in Paris deliberately.'
- b. Anna zit in Parijs om de Mona Lisa te zien.  
Anna sits in Paris in.order the Mona Lisa to see  
'Anna is in Paris to see the Mona Lisa.'
- c. Anna zit uit zichzelf in Parijs.  
Anna sits by herself in Paris  
'Anna is in Paris by herself.'

Even though the sentences are undeniably fully acceptable, it would be odd to conclude that a state has an agentive subject. This would go against everything we know about states: states have holders of states as subjects, not agents (Kratzer 1996). It would also contradict the mere meaning of the sentences. The sentences in (64) do not mean that Anna is actually doing something, she is just being somewhere. Now, as pointed out in Alexiadou, Anagnostopoulou and Schäfer (2015:20-21) agentivity tests in some contexts merely may react to the animacy and volition of the subject and general world knowledge, rather than truly indicating syntactic agentivity (see Marelj 2004; Kallulli 2007, Schäfer 2008, 2009; Koontz-Garboden 2009). Indeed, when the subject is inanimate, the tests fail:

- 65) a. \* Er zit opzettelijk suikerin de koffie.<sup>117</sup>  
there sits deliberately sugar in the coffee

---

<sup>117</sup> This sentence is acceptable if someone deliberately put sugar in the coffee. However, the sugar itself cannot have the intention to be in the coffee deliberately.

- b. \* Er zit suiker<sub>i</sub> in de koffie PRO<sub>i</sub> om te zoeten.  
 there sits sugar in the coffeePRO in.order to sweeten
- c. \* De suiker zit uit zichzelf in de koffie.  
 the sugar sits by itself in the coffee

Notice further that the same problem arises with the *bona fide* copula *zijn* 'to be'. Its locational readings are trusted to be states (Dowty 1979: 184, see also Filip 1999: 31-32). As such, they resist durative readings:

- 66) a. Jan is in Frankrijk.  
 John is in France  
 'John is in France.'
- b. \* Jan is in Frankrijk aan het zijn.  
 John is in France on the be.INFINITIVE

Yet, when its subject is animate, it passes agentivity tests:

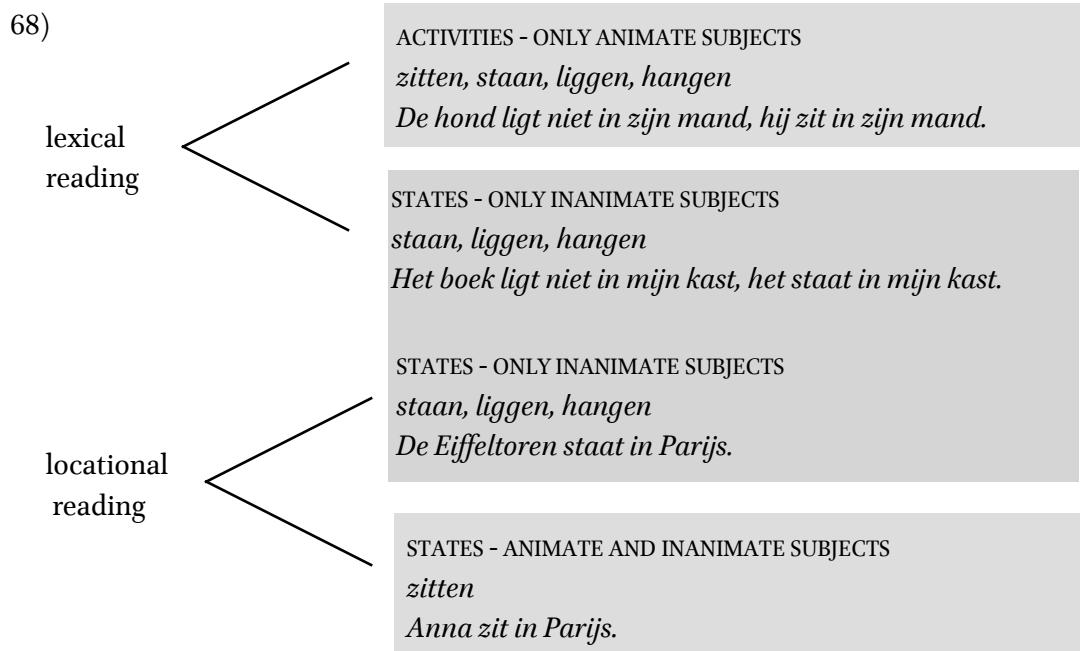
- 67) a. Anna is opzettelijk in Parijs.  
 Anna is deliberately in Paris  
 'Anna is in Paris deliberately.'
- b. Anna is in Parijs om de Mona Lisa te zien.  
 Anna is in Paris in.order the Mona Lisa to see  
 'Anna is in Paris to see the Mona Lisa.'
- c. Anna is uit zichzelf in Parijs.  
 Anna is by herself in Paris  
 'Anna is in Paris by herself.'

To the best of my knowledge, no one would claim that the locational copula construction in (67) is actually syntactically agentive, expressing an activity. I conclude that the durative reading is a much better indicator of *Aktionsart* than the agentivity tests. The agentivity tests seem to be in these cases merely sensitive to the animacy of the subject. The locational readings with *zitten* 'to sit' thus are states, just as the other locational readings, the subject of which is the holder of the state.

From the present section we may conclude that locational readings with *zitten* 'to sit' are states, just as other locational readings with posture verbs. Importantly, however, we have noticed a crucial difference between *zitten* and the other posture verbs in the locational reading: *zitten* 'to sit' can have animate subjects in this reading, unlike the other posture verbs. I have observed that *zitten* actually behaves on a par with the copula *zijn* 'to be' in this respect. This is an exciting observation: it may be the case that *zitten* 'to sit' actually is a true copula as well. In section 5 I will argue that this is indeed the case.

#### 6.4.5 Interim summary: the syntactic classification of posture verbs

In this section we have seen that posture verbs in lexical readings may function as activities or as states. In locational readings, they are always states. As such, the syntactic classification of the occurrences of these posture verbs intersects the distinction between the lexical and locational reading. The locational reading of the verb *zitten* ‘to sit’ patterns with the other locational readings, yet the lack of animacy restrictions on the subject indicates it should be treated separately. This is schematised in (68), in which the branches indicate the semantic distinction between the lexical and locational readings and the grey fields indicate the syntactic classification:



What distinguishes the activity reading from the stative reading is the presence or absence of syntactic heads. This indicates that these heads do not inherently belong to the verb and are thus not projected by the verb. What distinguishes the *lexical* state reading from the locational one, is the pragmatic context and the size of the ground, as argued in section 3.2. As such, these distinctions in the classification are of a very different, non-syntactic, perhaps pragmatic nature.

#### 6.4.6 The *Aktionsart* of motion verbs

The lexical readings of the motion verbs come in just one version. Being unaccusatives, the verbs *komen* ‘to come’ and *vallen* ‘to fall’ lack a Voice head in general and they are thus always non-agentive (see Alexiadou, Anagnostopoulou, Schäfer 2015).<sup>118</sup> However, given that they are accomplishments aspectually, durative readings are possible under the lexical reading:

<sup>118</sup> For *vallen* ‘to fall’ I can get a reading in which someone is falling deliberately (e.g. a stuntman or an insurance fraud). In such contexts, *vallen* is arguably used in an agentive activity reading. Given that a discussion of the *aktionsart* of lexical *vallen* ‘to fall’ is not essential to the discussion, I ignore these facts.

- 69) a. De komeet was aan het **vallen**.  
          the comet was on the fall.INFINITIVE  
          'The comet was falling.'
- b. Ik voel dat de veranderingaan het **komen** is.  
          I feel that the change on the come.INF is.  
          'I feel change is coming.'

I conclude that the lexical reading is non-agentive and lacks a Voice head, but does contain an unbounded little v. Under the locational reading -which for the motion verbs should more accurately be called the temporal and the origin reading- such a durational reading is excluded:

- 70) \* Die oorlog was in de negentiende eeuw aan het **vallen**.  
          that war was in the nineteenth century on the fall.INFINITIVE
- 71) \* Mijn vrouw is uit Frankrijk aan het **komen**.  
          my wife is from France on the come.INFINITIVE  
          (origin reading intended, i.e. 'I feel my wife is French.')

I conclude that the origin reading of *komen* and the temporal reading of *vallen* are states: they contain a stative little v and no Voice head.

Note that *komen* 'to come' can get both animate and inanimate subjects:

- 72) a. Jean komt uit Frankrijk.  
          Jean comes from France  
          'Jean is from France.'
- b. Camembert komt uit Frankrijk.  
          Camembert comes from France  
          'Camembert is from France.'

As such, it behaves as *zitten* 'to sit' in locational readings. However, unlike *zitten* it does not pass agentivity tests:

- 73) \* Jean komt opzettelijk uit Frankrijk.<sup>119</sup>  
          Jean comes deliberately from France  
          Intended: \* 'Jean is deliberately from France, i.e. he is deliberately French.'

Now note that the copula *zijn* 'to be' does not pass agentivity tests in this context either:

---

<sup>119</sup> I can imagine a grammatical reading for the sentence. For example, it is possible in a context in which an author states that the character *Jean* in the novel was intentionally chosen to be French. The adverb then refers to the intentions of the author, not to the intentions of *Jean*.

- 74) \* Jean is opzettelijk uit Frankrijk.<sup>120</sup>  
 Jean is deliberately from France  
 Intended: \* 'Jean is deliberately from France.'

The contrasting behaviour of agentivity for *zitten* and *komen* in this context and the fact that the copula *zijn* 'to be' shows exactly the same contrasts shows once again that the agentivity tests may correspond with world knowledge about volitionality rather than true syntactic agentivity in some contexts.

The verb *vallen* 'to fall' invariably requires event nouns as subjects in temporal readings. To see this, first note, for good measure, that *feest* 'party' and *val* 'fall' are event nouns, in contrast with, for example, *stoel* 'chair' which refers to an object. As such, *feest* and *val* can be modified for duration, whereas *stoel* cannot:

- 75) a. \* een stoel van drie uur  
           a chair of three hours
- b. een feest van drie uur  
           a party of three hours  
           'a party that takes three hours'
- c. een val van drie uur  
           a fall of three hours  
           'a fall that takes three hours'

Furthermore, the Dutch preposition *voor* 'before/in front of' will be interpreted temporally for *feest* and *val*, but spatially for *stoel*:

- 76) a. voor de stoel      ✓location/\*moment in time  
       before the chair  
       'in front of the chair'
- b. voor het feest      ??location/✓moment in time  
       before the party  
       'before the party'
- c. voor de val      ??location/✓moment in time  
       before the fall  
       'before the fall'

Now consider that only the event nouns are licit with the temporal reading of *vallen* 'to fall':

---

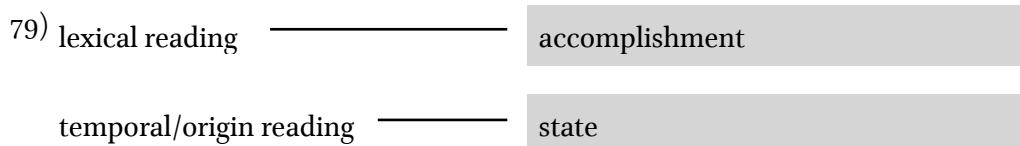
<sup>120</sup> The comment in the previous footnote holds for this example as well.

- 77) a. # De stoel viel in de nacht.  
          the chair fell in the night  
       Lexical reading: 'The chair fell down at night.'  
       \* Temporal reading: 'The chair took place at night.'
- b. Het feest viel in de namiddag.  
      the party fell in the afternoon  
      'The party was in the afternoon.'
- c. De val van de muur viel in de 20<sup>ste</sup> eeuw.  
      the fall of the wall fell in the 20<sup>th</sup> century  
      'The fall of the wall was in the 20<sup>th</sup> century.'

The copula *zijn* 'to be' shows the same restriction in temporal contexts:

- 78) a. \* De stoel was in de nacht.  
          the chair was in the night
- b. Het feest was in de namiddag.  
      the party was in the afternoon  
      'The party was in the afternoon.'

To conclude, for the motion verbs, the syntactic distinction and the semantic distinction are therefore aligned, as shown in the following scheme:



Furthermore, we have seen that *vallen* in the temporal reading and *komen* in the origin reading show the same restrictions on subjects as *zijn* 'to be' does in similar contexts. Again, given that they behave on a par with *zijn* 'to be', one may suspect they are copulas as well. In the following sections I will argue this is indeed the case.

#### 6.4.7 Locational readings and *zijn* 'to be'

We are now in a position to understand why the locational readings of the posture verbs and the temporal and origin reading of the motion verbs intuitively feel so close to being copulas. The *bona fide* stative copula *zijn* 'to be' is also a stative verb. It shows exactly the same aspectual behaviour in the context of a locational predicate, a temporal predicate and a directional predicate expressing origin:

- 80) a. Jan is in Frankrijk.  
          John is in France  
       'John is in France.'

- b. \* Jan is in Frankrijk aan het zijn.  
 John is in France on the be.INFINITIVE

- 81) a. De MonaLisa is in het Moma voor een tentoonstelling.  
 the Mona Lisa is in the MoMa for an exhibition  
 'The Mona Lisa is at the MoMa for an exhibition.'

- b. \* De MonaLisa is in het Moma aan het zijn.  
 the Mona Lisa is in the MoMa on the be.INF

- 82) a. Die oorlog was in de negentiende eeuw.  
 that war was in the nineteenth century  
 'That war was in the nineteenth century.'

- b. \* Die oorlog was in de negentiende eeuw aan het zijn.  
 that war was in the nineteenth century on the be.INFINITIVE

- 83) a. Jean is uit Frankrijk.  
 Jean is from France  
 'Jean is from France.'

- b. \* Jean is uit Frankrijk aan het zijn.  
 Jean is from France on the be.INFINITIVE

- 84) a. Parijs is in Frankrijk.  
 Paris is in France  
 'Paris is in France.'

- b. \* Parijs is in Frankrijk aan het zijn.  
 Paris is in France on the be.INFINITIVE

The locational reading of the posture verbs and the temporal and origin readings of the motion verbs have their stativeness in common with the copula. In short, when it comes to the stative-durative opposition, the locational readings and the non-durative lexical readings pattern with the verb *zijn* 'to be' rather than with their own durative, lexical readings.

#### 6.4.8 Conclusion

In this section we have seen that the functional superstructure of posture verbs and motion verbs can alternate between activities and states. We have ascribed this contrast to the presence or absence of functional heads in the verbal spine. This syntactic classification intersects the lexical-locational readings.

Importantly, we have seen that the locational reading of the posture verbs and the temporal and origin readings of the motion verbs have their stativeness in common with the copula *zijn* 'to be' in

locational readings.<sup>121</sup> If we then take into account the fact that the prepositional phrases of the posture verbs and the motion verbs can be interpreted easily as predicates of copulas, it should not come as a surprise that these locational readings *feel* like copula readings.

However, even though we have seen that all posture and motion verbs acquire a copula-like *Aktionsart* and meaning in locational/temporal/origin readings, there are only three verbs that exactly show the same restrictions on subjects as *zijn* 'to be' in the relevant contexts. These verbs are *zitten* 'to sit', *vallen* 'to fall' and *komen* 'to come'. If Dutch were a splitting languages, these are the three verbs we should be looking out for. Is it possible that they truly *are* copulas in the locational/temporal/origin readings? The next two sections address this question. I will conclude that, indeed, they are copulas.

## 6.5 Copulas

Copular sentences are expanded small clauses (Stowell 1978 and see also Moro 1997:chapter 1 a.o.). As a result, there is a tight relation between copular sentences and small clauses: a small clause plus a copula is a copular sentence, a copular sentence without a copula is a small clause. An empirical prediction that follows is that small clauses should not contain copulas, by definition. The following example illustrates this for the *bona fide* default Dutch copula *zijn* 'to be'. In a full clause, the copula is obligatorily overtly realised, in small clauses it is obligatorily absent:

- 85) a. Anna is lief.  
Anna is kind  
'Anna is kind.'
- b. Ik vind Anna lief.  
I consider Anna kind  
'I consider Anna kind.'
- c. \* Ik vind Anna lief zijn.  
I consider Anna kind to.be

For copular clauses with a PP, the effect can be observed with small clauses that are the complement of sentential verbs:

- 86) a. Anna is in de tuin.  
Anna is in the garden  
'Anna is in the garden.'
- b. Ik zag Anna in de tuin.  
I saw Anna in the garden  
'I saw Anna in the garden.' (Intended: I saw that Anna was in the garden.)

---

<sup>121</sup> In Dowty's (1979) semantic terms, they are all momentary states.

- c. \*? Ik zag Anna in de tuin zijn.  
 I saw Anna in the garden be  
 Intended: 'I saw Anna in the garden.'

In this respect copulas contrast with lexical verbs. They can be present in similar contexts:

- 87) a. Anna werkt in de tuin.  
 Anna works in the garden  
 'Anna is in the garden.'
- b. Ik zag Anna in de tuin.  
 I saw Anna in the garden  
 \* 'I saw Anna working in the garden.'
- c. Ik zag Anna in de tuin werken.  
 I saw Anna in the garden work  
 'I saw Anna working in the garden.'

The prediction is now straightforward: those verbs that *can* occur overtly in such small clauses under the locational reading are lexical verbs, those that cannot occur in such small clauses are copulas. I will now test the Dutch posture verbs and motion verbs one by one.

Given the conclusion of the previous section we suspect that *zitten*, *vallen* and *komen* might be true copulas. If this is borne out, they should be illicit in small clauses. For the posture verbs the grounds are spacious grounds in the examples below to force the desired locational reading. To improve the pragmatic plausibility of the following sentences, imagine that they are all responses to the question 'what did you see in your dreams tonight?' As the following examples show, this is borne out.

- 88) # Ik zag Anna in Frankrijk zitten.  
 I saw Anna in France sit  
 'I saw Anna sitting in France.' (lexical reading, pragmatically odd)  
 \* 'I saw Anna in France.' (locational reading, excluded)
- 89) \*? Ik zag de tweede wereldoorlog in de middeleeuwen vallen.  
 I saw the second world.war in the Middle.ages fall  
 Intended: 'I saw the second world war taking place in te Middle Ages.'
- 90) # Ik zag een Franse president uit Spanje komen.  
 I saw a French president from Spain come  
 'I saw the French president coming/arriving from Spain.' (lexical reading)  
 \* 'I saw a French president who was from Spain.'

The examples indicate that *zitten*, *vallen* and *komen* can function as true copulas, as they are illicit in the small clauses under the intended locational readings.

We further expect *staan* 'to stand', *liggen* 'to lie' and *hangen* 'to hang' not to be copulas and thus to appear in the small clauses. Indeed, this is borne out:

- 91)    Ik zag de Eiffeltoren in Spanje staan.  
I saw the Eiffel.tower in Spain stand  
'I saw the Eiffel Tower standing in Spain.'
- 92)    Ik zag de Mona Lisa in Spanje hangen.  
I saw the Mona Lisa in Spain hang  
'I saw the Mona Lisa hanging in Spain.'
- 93)    Ik zag de Statenbijbel in Spanje liggen.  
I saw the States.bible in Spain lie  
'I saw the States bible lying in Spain.'

The following examples prove the same point: *zitten* contrasts with *staan*, *liggen* and *hangen*. They contain small clauses in which the prepositional phrase is omitted. The grammaticality contrasts become even sharper:

- 94)    #    Ik zag Anna zitten, maar dan in Frankrijk.  
I saw Anna sit,       but then in France  
'I saw Anna sitting, but in France.' (lexical reading, pragmatically odd)  
\* 'I saw Anna, but she was in France.' (locational reading, excluded)
- 95)    \*    Ik zag de tweede wereldoorlog vallen, maar dan in de middeleeuwen.  
I saw the second world.war fall,       but then in the Middle.ages  
Intended: 'I saw the second world war taking place in te Middle Ages.'
- 96)    \*    Ik zag een Franse president komen, maar dan uit Spanje.  
I saw a French president come,       but then from Spain  
Intended: 'I saw a French president, but he was born in Spain.'
- 97)    Ik zag de Eiffeltoren staan, maar dan in Spanje.  
I saw the Eiffel.tower stand, but then in Spain  
'I saw the Eiffel Tower, but it was in Spain.'
- 98)    Ik zag de Mona Lisa hangen, maar dan in Spanje.  
I saw the Mona Lisa hang,       but then in Spain  
'I saw the Mona Lisa hanging, but it was in Spain.'

- 99)       Ik zag de Statenbijbel liggen, maar dan in Spanje.  
          I saw the States.bible lie,       but then in Spain  
          'I saw the States Bible lying, but it was in Spain.'

Note that the illicit sentences improve tremendously, as soon as we omit the verb from the small clause:

- 100)      Ik zag Anna, maar dan in Frankrijk.  
          I saw Anna sit, but then in France  
          'I saw Anna, but she was in France.'
- 101)      Ik zag de tweede wereldoorlog, maar dan in de middeleeuwen.  
          I saw the second world.war     but then in the Middle.Ages  
          'I saw the second world war, but it was during the Middle Ages.'
- 102)      Ik zag een Franse president, maar dan uit Spanje.  
          I saw a French president, but then from Spain  
          'I saw a French president, but he was from Spain.'

These facts indicate that *zitten* 'to sit', *vallen* 'to fall' and *komen* 'to come' are copulas in Dutch. They each have their own designated semantics: *zitten* 'to sit' expresses a location, *vallen* 'to fall' has a temporal reading and *komen* 'to come' refers to an origin. These verbs must have been grammaticalised into copulas. When they are used in a locational/temporal/origin reading, they thus do not simply show malleable behaviour. These verbs differ from the location verbs *staan* 'to stand', *liggen* 'to lie' and *hangen* 'to hang', which did not grammaticalise into copulas.

We can conclude that Dutch is indeed a splitting language typologically as it has more than one stative copula. The next section is a cross-linguistic comparison which supports the proposed bifurcation into copulas and non-copulas. In later sections I will return to the Dutch copulas and I will discuss their semantics, their syntax and the grammaticalisation in more detail.

## 6.6 Posture verbs and motion verbs in continental West-Germanic varieties

### 6.6.1 Available locational readings

A cross-linguistic comparison indicates that the distinction between copulas and non-copulas in the domain of posture verbs and motion verbs is justified. If we compare the Dutch data with other continental West-Germanic data from Frisian, Low German and High German (i.e. Standard German), it becomes clear that the posture verbs and motion verbs do not behave as a uniform group in these languages, but that they show the same bifurcation: their posture verbs and motion verbs can be copulas or non-copulas. More specifically, the generalisation will be as follows: the non-copular verbs can have the locational readings in all varieties. The copular verbs depend on

grammaticalisation to be licit in a locational reading and whether they have undergone such grammaticalisation is language dependent.

Let us first take a look at the set of verbs that behave alike in all these continental West-Germanic varieties in the sense that they do occur in locational constructions.<sup>122</sup> The examples from (103) to (106) present the data for the verbs for ‘to stand’, ‘to lie’, ‘to hang’ and ‘to come’. For reasons of brevity, I gloss them in series when possible.

- |      |   |
|------|---|
| 103) | <ul style="list-style-type: none"> <li>✓ De Eiffeltoren staat in Frankrijk. [Dutch]</li> <li>✓ De Eiffeltoer stjit yn Frankryk. [Frisian]</li> <li>✓ Dei Eiffelturm stait in Frankriek. [Low German]</li> <li>✓ Der Eiffelturm steht in Frankreich. [High German]</li> </ul> <p>the Eiffel.tower stands in France<br/>         ‘The Eiffel Tower is in France.’</p> |
| 104) | <ul style="list-style-type: none"> <li>✓ Normandië ligt in Frankrijk. [Dutch]</li> <li>✓ Normandië leit yn Frankryk. [Frisian]</li> <li>✓ Normandie liggt in Frankriek. [Low German]</li> <li>✓ Die Normandie liegt in Frankreich. [High German]</li> </ul> <p>the Normandy lies in France<br/>         ‘Normandy is in France.’</p>                                |
| 105) | <ul style="list-style-type: none"> <li>✓ De Mona Lisa hangt in het Louvre. [Dutch]</li> <li>✓ De Mona Lisa hinget yn ‘t Louvre. [Frisian]</li> <li>✓ Dei Mona Lisa hangt in d’ Louvre. [Low German]</li> <li>✓ Die Mona Lisa hängt i- m Louvre. [High German]</li> </ul> <p>the Mona Lisa hangs in the Louvre<br/>         ‘The Mona Lisa is in the Louvre.’</p>    |
| 106) | <ul style="list-style-type: none"> <li>✓ Ze komt uit Frankrijk. [Dutch]</li> <li>✓ Ze komt út Frankryk. [Frisian]</li> <li>✓ Zei kumt ut Frankriek. [Low German]</li> <li>✓ Sie kommt aus Frankreich. [High German]</li> </ul> <p>she comes from France<br/>         ‘She is from France.’</p>  |

For the verbs *zitten/sitten/sitzen* ‘to sit’ there is variation: it behaves as it does in Dutch in Frisian and in Low German, but not in High German. The contrast is strong: the High German examples are strongly ungrammatical, whereas the examples are fully acceptable in the other varieties.

---

<sup>122</sup> The Low German data that are reported in this section are Low German as spoken in the Krummhörn region, which is situated in the Northwest of Germany. It is separated from the Dutch border by the Ems estuary. I would like to thank (anonymised) for help with these data. I would like to thank (anonymised) for help with the Frisian data.

- 107) ✓ Hij zit in Frankrijk. [Dutch]  
 ✓ Hie zit yn Frankryk. [Frisian]  
 ✓ Hai sit in Frankriek. [Low German]  
 \* Er sitzt in Frankreich. [High German]  
 he sits in France  
 'He is in France.'
- 108) ✓ Er zit suiker in de koffie. [Dutch]  
 ✓ Der sit sūker yn 'e kofje [Frisian]  
 ✓ Doar sit Zucker in d' Koffje. [Low German]  
 \* Es sitzt Zucker i- m Kaffee. [High German]  
 there sits sugar in the coffee  
 'There is sugar in the coffee.'

Finally, the acceptability of *vallen/fallen* ‘to fall’ in temporal constructions is a bit more diffuse. The following sentence is certainly licit in Dutch, but not in the other West-Germanic languages. Speakers of High German reject the example after some hesitation, but they do reject it. One might conjecture that the High German verb *fallen* ‘to fall’ is on its way to grammaticalise into a copula, as in Dutch, but it certainly did not reach full acceptability yet.

- 109) ✓ De oorlog viel in de twintigste eeuw. [Dutch]  
 \*? De oarloch foel yn 'e tweintichste iuw. [Frisian]  
 \*? Dei Krieg ful in d' twintigste Joarhunnert. [Low German]  
 \*? Der Krieg fiel in-s zwanzigsten Jahrhundert. [High German]  
 the war fell in the twentieth century  
 'The war was in the twentieth century.'

Examples that belong to a lower register were rejected immediately and it was suggested that the default copula (the language’s version of ‘to be’) should be used instead. Again, the Dutch example is completely fine.

- 110) ✓ De kerkdienst viel vroeg in de morgen. [Dutch]  
 \* Kark ful frau in d' Morgen. [Low German]  
 the church.service fell early in the morning  
 'The church service was early in the morning.'
- \* De fiering foel moarns betiid [Frisian]  
 the celebration fell mornings early
- \* Der Gottesdienst fiel am frühen Morgen. [High German]  
 the church.service fell in.the early morning

Note that the German example is fine with the copula *sein* 'to be', so the ungrammaticality of the sentence is to be ascribed to the fact that *fallen* 'to fall' cannot serve as a copula in this sentence:

- 111) ✓ Der Gottesdienst war am frühen Morgen. [High German]  
 the church.service was in.the early morning  
 'The church service was early in the morning.'

What is accepted in all varieties, is the following construction, which has the following two properties: there is a fixed preposition that translates into English as 'on' and there is some sense of a repeating event of which the timing varies. I do not understand why exactly this construction is licit in all varieties, in contrast with the examples discussed above. Note that the German example above becomes acceptable if the preposition is changed into *auf* 'on'. I conjecture that the verb *vallen/fallen* forms a lexicalised idiomatic collocation together with the preposition in this sentence. I mention these data for good measure, but I have nothing further to say about them.

- 112) ✓ De eerste mei viel op een dinsdag. [Dutch]  
 ✓ De earste maaie foel op in tiisdei [Frisian]  
 ✓ Dei eerste Mai ful op en Diensdach [Low German]  
 ✓ Der erste Mai fiel auf einen Dienstag. [High German]  
 the first May fell on a Tuesday  
 'The first of May fell on a Tuesday.'
- 113) ✓ Der Gottesdienst fiel auf den frühen Morgen. [High German]  
 the church service fell in the early morning  
 'The church service took place in the early morning.'

The observations are summarised in the following table:

114)

	Dutch	Frisian	Low German	High German
to sit	✓	✓	✓	✗
to stand	✓	✓	✓	✓
to lie	✓	✓	✓	✓
to hang	✓	✓	✓	✓
to come	✓	✓	✓	✓
to fall	✓	✗	✗	✗

Considering the table, it becomes clear that the mere observation that posture and motion verbs can occur with a locational meaning rather than with a full lexical meaning does not suffice to capture the cross-linguistic variation. It seems that something special needs to be said about at least *zitten/sitten/sitzen* ‘to sit’ and *vallen/fallen* ‘to fall’. Recall that I argue that Dutch *zitten*, *komen* and *vallen* have grammaticalised into true copulas in Dutch. The table now suggests that these verbs can only occur in the locational/temporal/origin readings in a language variety if and only if they have grammaticalised into copulas in that variety. We then predict the following copula inventories:

- 115) Frisian: *sitte* ‘to sit’, *komme* ‘to come’  
 Low German: *sitten* ‘to sit’, *komen* ‘to come’  
 High German: *kommen* ‘to come’

In the following subsections, I present the small clause tests for these varieties and we will see that the predictions are borne out.

### 6.6.2 Copulas in Frisian

For the Frisian verbs for ‘to stand’, ‘to lie’ and ‘to hang’, we predict that they survive in small clauses as they are argued to be lexical verbs. This prediction is borne out. Again, imagine the sentences are replies to the question ‘What did you see in your dreams last night?’

- 116) Ik seach de Eiffeltoer yn Spanje stean.  
 I saw the Eiffel.tower in Spain stand  
 ‘I saw the Eiffel Tower in Spain.’

- 117) Ik seach de Mona Lisa yn Spanje hingjen.  
 I saw the Mona Lisa in Spain hang  
 ‘I saw the Mona Lisa hanging in Spain.’

- 118)    Ik seach de bybel yn Spanje lizzen.  
       I saw the bible in Spain lie  
       'I saw the bible lying in Spain.'

In contrast, for the verbs for 'to sit' and 'to come', we expect that they cannot occur in small clauses. Again, this is borne out, as in Dutch:

- 119)    # Ik seach Anna yn Frankryk sitte.  
       I saw Anna in France sit  
       'I saw Anna sitting in France.' (lexical reading, pragmatically odd)  
       \* 'I saw Anna in France.' (locational reading, excluded)
- 120)    \* Ik seach in Frânske president út Spanje kommen.  
       I saw a French president from Spain come  
       'I saw a French president who was from Spain, i.e. who was Spanish.'

Finally and trivially, given that the temporal reading for 'to fall' is excluded anyway in Frisian, it cannot occur in a small clause either:

- 121)    \*? Ik seach de twadde wrâldoorloch yn de Midsieuwen fallen.  
       I saw the second world.war in the Middle.ages fall  
       Intended: 'I saw the second world war taking place in the Middle Ages.'

I conclude that Frisian is a splitting language, just like Dutch. However, it has grammaticalised only the verbs for 'to sit' and 'to come' into stative copulas. The verb for 'to fall' has not been grammaticalised into a copula and, consequently, is illicit under a temporal reading. As such, it contrasts with the verbs for 'to stand', 'to hang' and 'to lie', which do not depend on a copular status to occur in a locational reading.

### 6.6.3 Copulas in Low German

The verbs show exactly the same behaviour in Low German as they did in Frisian:

- 122)    Ik sach dei Eiffelturm in Spanien stoan.  
       I saw the Eiffel.tower in Spain stand  
       'I saw the Eiffel Tower in Spain.'
- 123)    Ik sach dei Mona Lisa in Spanien hangn.  
       I saw the Mona Lisa in Spain hang  
       'I saw the Mona Lisa hanging in Spain.'
- 124)    Ik sach dei Bibel in Spanien liergn.  
       I saw the bible in Spain lie  
       'I saw the States bible lying in Spain.'

- 125) # Ik sach en Französische Präsident ut Spanien komm.  
       I saw a French president from Spain come  
       \*I saw a French president who was from Spain.' (copula reading)  
       'I saw a French president arriving from Spain.' (lexical reading)
- 126) \* Ik sach dei twaid Weltkrieg in d' Mittelalter falln.  
       I saw the second world.war in.the Middle.ages fall  
       Intended: 'I saw the second world war taking place in the Middle Ages.'
- 127) # Ik sach Anna in Frankriek sittn.  
       I saw Anna in France sit  
       'I saw Anna sitting in France.' (lexical reading, pragmatically odd)  
       \* 'I saw Anna in France.' (locational reading, excluded)

I conclude that Low German has grammaticalised the same two verbs into copulas, i.e. the verbs for 'to come' and 'to sit'. The verb for 'to fall' cannot occur in a locational reading as, apparently, the licitness of this verb in this reading depends on copular status.

#### 6.6.4 Copulas in High German

We predict that in High German, the verbs for *to stand*, *to hang* and *to lie* are lexical verbs and thus occur overtly in small clauses. This is borne out:

- 128) Ich sah den Eiffelturm in Spanien stehen.  
       I saw the Eiffel.tower in Spain stand  
       'I saw the Eiffel Tower in Spain.'
- 129) Ich sah die Mona Lisa in Spanien hängen.  
       I saw the Mona Lisa in Spain hang  
       'I saw the Mona Lisa hanging in Spain.'
- 130) Ich sah die Bibel in Spanien liegen.  
       I saw the bible in Spain lie  
       'I saw the States bible lying in Spain.'

The verb *kommen* 'to come' does only have its lexical reading and not its copular reading, as predicted:

- 131) # Ich sah ein Französisch Präsident aus Spanien kommen.  
       I saw a French president from Spain come  
       'I saw a French president arriving from Spain.'  
       \* 'I saw a French president who was from Spain.'

As expected, the verbs *fallen* 'to fall' and *sitzen* 'to sit' are not licit in the small clause under the copula reading, but they were not acceptable under the locational reading in the main clause either,

so these examples tell us nothing beyond what we already knew: these verbs do not function as copulas in High German.

- 132) \* Ich sah den zweiten Weltkrieg im Mittelalter fallen.  
I saw the second world.war in.the Middle.ages fall  
Intended: 'I saw the second world war taking place in the Middle Ages.'

- 133) # Ich sah Anna in Frankreich sitzen.  
I saw Anna in France sit  
'I saw Anna sitting in France.' (lexical reading, pragmatically odd)  
\* 'I saw Anna in France.' (locational reading, excluded)

In short, of the verbs under discussion only *kommen* 'to come' grammaticalised into a copula in High German.

### 6.6.5 Conclusion

We have seen that in all continental West-Germanic varieties under discussion, the verbs for 'to stand', 'to lie' and 'to hang' have a locational reading and in all varieties they are not copulas in this context. In all varieties, the verbs for 'to sit', to fall' and 'to come' either do not occur in a locational/temporal/origin reading or when they occur, they are copulas. As such, we can conclude that the licitness of these verbs in the locational readings depends on copular status. In Dutch, these three verbs have all acquired a copular status. In High German, *kommen* 'to come' is a copula as well. *Sitzen* 'to sit' and *fallen* 'to fall' are rejected as copulas. Frisian and Low German accept the verbs for 'to sit' and 'to come' as copulas, but not 'to fall.'

## 6.7 The semantics of the posture and motion copulas

### 6.7.1 Introduction

This section discusses the semantics of the copulas under discussion. It will become clear that the posture verb *zitten* 'to sit' and the motion verbs *vallen* 'to fall' and *komen* 'to come' are nothing alike semantically. In the following sections, we will see that this is so for a principled reason. I first describe the facts in the following two subsections. As posture verbs and motion verbs have distinct properties, *zitten* and *vallen/komen* are discussed in two separate subsections. Later subsections then present an account for the facts.

### 6.7.2 The semantics of zitten as a copula

When the posture verb *zitten* ‘to sit’ is used as a copula the predicate in the sentence is always locational. Syntactically, this predicate is typically a prepositional phrase, such as *in Frankrijk* ‘in France’, but adverbial predicates, such as *daar* ‘there’ are possible as well:

- 134) Zit je vrouw nog altijd **in Frankrijk?** Ja, ze zit daar al lang.  
sits your wife still always in France yes she sits there already long  
'Is your wife still in France? Yes, she has been there for quite a while now.'

Even though the predicate is restricted to locations, what counts as a location is quite liberal. Metaphoric and idiomatic prepositional phrases are certainly not excluded. In the following examples, the location is interpreted idiomatically:

- 135) a. Ze zit **op haar limiet.**  
she sits on her limit  
'She can't take it anymore.'
- b. Ze zit **in de put.**  
she sits in the hole  
'She is depressed.'
- c. Ze zit **in mijn hart.**  
she sits in my heart  
'I love her.'

However, it is still important that the predicate is interpretable as a location, albeit metaphorically. All other kind of predicates, such as nominal or adjectival predicates are strongly excluded:

- 136) a. Ze is/\*zit een vrouw. \*NP complement  
she is/\*sits a woman  
Intended: 'She is a woman.'
- b. Ze is/\*zit intelligent. \*AP complement: individual-level predicate  
she is/\*sits intelligent  
Intended: 'She is intelligent.'
- c. Ze is/\*zit dronken. \*AP complement: stage-level predicate  
she is/\*sits drunk  
Intended: 'She is drunk.'
- d. Ze is/\*zit stervende. \*complement is a participle  
she is/\*sits dying  
Intended: 'She is dying.'

- e. Wees/\*zit jezelf. \*complement is a pronoun  
 be/\*sit yourself  
 Intended: 'Be yourself.'

Also excluded as predicates are directional predicates. Every predicate that refers to a path is illicit (see also Zagona 2012 on Spanish and see below):

- 137) a. \* Ze zit naar Frankrijk.  
           she sit to France  
 b. \* Ze zit uit Frankrijk.  
       she sits from France

In the example above, *naar Frankrijk* 'towards France' and *uit Frankrijk* 'from France' both refer to paths. The examples are fully illicit. I will come back to this restriction.

A further restriction on *zitten* 'to sit' as a copula is that it does not tolerate event nouns as subjects (again, see also Zagona 2012 on Spanish and see below). Recall from section 4.6 that *feest* 'party' and *val* 'fall' are event nouns. As the subjects of the copula *zitten*, they are illicit:

- 138) a. Het feestje is/\*zit in de disco. subject is an event noun  
       the party is/\*sits in the disco  
       Intended: 'The party takes place in the disco.'
- b. De val van de muur was/\*zat in Berlijn. subject is an event noun  
       the fall of the wall was/\*sat in Berlin  
       Intended: 'The fall of the wall happened in Berlin.'

Note, for good measure, that, similarly, an event noun is illicit as the predicate of the copula *zitten*. Of course, we have seen that event nouns are interpreted temporally and that posture verbs require a locational predicate, so this restriction is perhaps less surprising than the observation that an event noun is illicit as a subject in this context.

- 139) a. \* Een open grens zat bij de val van de muur. predicate is an event noun  
       an open border sat at the fall of the wall
- b. \* Het feest zat al tijdens de val. both are event nouns  
       the party sat already during the fall

In sum, in this section I have discussed restrictions on the use of the posture verb *zitten* 'to sit' as a copula. We have seen that the predicate needs to refer to a location, literally or metaphorically. As such, the predicates are prepositional phrases or adverbs. We have further seen that directional predicates are excluded: *zitten* 'to sit' is incompatible with paths. Finally, we have seen that event nouns do not combine with the copula *zitten* 'to sit' either, neither as subjects or as predicates.

### 6.7.3 The semantics of the motion verbs as copulas

The motion verbs that can be used as copulas are *komen* ‘to come’ and *vallen* ‘to fall’. *Komen* can only co-occur with predicates that are paths that refer to an origin (and see Zagona 2012 on Spanish, to which I come back below):

- 140) Ze komt uit het buitenland.  
she comes from the abroad  
'She is from abroad.'
- 141) Ze komt van een goede familie.  
she comes of a good family  
'She is from a respectable family.'

Note that *komen* contrasts sharply with the copula *zitten* in this respect. Whereas *zitten* ‘to sit’ is incompatible with paths, *komen* ‘to come’ requires a path.

The copula *vallen* is restricted to temporal predicates, as shown in (142)-(143). Its subject is invariably an event noun, as I have discussed in section 4.6.

- 142) Die oorlog viel in de twintigste eeuw.  
that war fell in the twentieth century  
'That war was in the twentieth century.'
- 143) De les viel in de namiddag.  
the class fell in the afternoon  
'The class was in the afternoon.'

Again, we see that a motion verb obligatorily shows a property, i.e. an event noun as a subject, that was excluded for *zitten* ‘to sit’. In sum, we see that motion verbs require types of subjects and predicates that were excluded for *zitten* ‘to sit’, viz. paths as predicates and event nouns as subjects. The observations are summarised in the following table:

144)

copula	subject	predicate	type of verb
zitten	animate/inanimate	spatial: place	posture verb
komen	animate/inanimate	spatial: path	motion verb
vallen	event noun	temporal	motion verb

In the typology on copulas, Spanish is quite famous for being a splitter. In other words, Spanish employs a different copula for nominal predicates, i.e. *ser*, and location predicates, i.e. *estar*. The precise distribution of *ser* and *estar* is complex (see see Diesing 1988, Kratzer 1995, Maienborn 2005, Camacho 2012) and doing justice to the complexity of the facts and the many insights that have been formulated is well beyond the aim of the present article.

What does concern us here, however, is the following observation by Zagona (2012). Zagona (2012) observed that *estar*, the copula for locative predicates, is incompatible with paths and event nouns. In this respect, the Spanish locational copula *estar* behaves on a par with the Dutch copula *zitten* ‘to sit’. It appears that cross-linguistically, the following observation may hold: posture verbs that grammaticalise into copulas resist paths and event nouns.

## 6.8 The syntax of posture and motion copulas

Den Dikken (2006) argues that copulas are semantically empty functional heads. The literature that presents structural decompositions of the verbal spine presents a different view: there it is argued that copulas are light verbs that spell out eventive heads. For example, Ramchand (2008: 106) mentions that the copula *to be* may lexicalise the head she calls *initiator* (*init*). This head is the most minimal event head, which, she argues is the only head contained in stative events. Other scholars might call this head *v* or *v<sub>STATE</sub>*. The same idea can be found in Alexiadou, Anagnostopoulou and Schäfer (2015: 49–50), who argue that copulas are light verbs and thus spell out functional heads in the verbal spine. They propose, for example, that *become* realizes a bare *v<sub>CAUS</sub>*. More specifically, they conjecture that copulas always spell out an eventive head, i.e. a *little v*.

In proposals on the eventive head *v* it is often argued that there are two variants: an action/change-of-state event and a state event (see, for example Ramchand 2008 and see the discussion on Kratzer 1996 below). As such, one can immediately capture the fact that languages typically contain both stative copulas, such as *to be* in English, and change-of-state copulas, such as *to become* in English. The change-of-state events are not relevant for the discussion at hand, as all copulas under discussion here are stative. I will therefore focus on the stative copulas in what follows.<sup>123</sup>

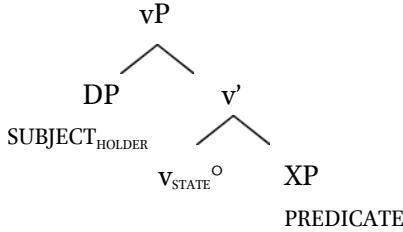
The present proposal adopts the insight that the structure of a stative copula contains an eventive head, *v<sub>STATE</sub>*, which introduces a stative event into the structure, linking a ‘holder’ of the state (in the terms of Kratzer 1996) to the state expressed by the predicate.

---

<sup>123</sup> In Standard Dutch the posture verbs and motion verbs happen to have acquired the stative *v*, but they could as well have acquired the change-of-state *v*. In fact, this is what happened in West-Flemish, a Dutch dialect family, for the motion verb *komen*. In West-Flemish, this verb functions as a change-of-state copula:

- 1) Ze is ziek aan het komen.  
she is ill on the come.INFINITIVE  
'She is becoming ill.'

145)



This  $v_{STATE}$  head is realised by the copula:

146)       $\text{zijn} \leftrightarrow [\text{v}_{\text{STATE}}]$ 

We now have determined what the default syntax is of a stative copula.

The continental West-Germanic languages behave as splitters: they have developed specific copulas for specific prepositional predicates. The Vocabulary Items are shown in (147):

147)       $\begin{array}{lll} \text{zijn} & \leftrightarrow [\text{v}_{\text{STATE}}] \\ \text{zitten} & \leftrightarrow [\text{v}_{\text{STATE}}] / \_\_\_ \text{PP}_{\text{LOCATION}} \\ \text{komen} & \leftrightarrow [\text{v}_{\text{STATE}}] / \_\_\_ \text{PP}_{\text{DIRECTION(OIGIN)}} \\ \text{vallen} & \leftrightarrow [\text{v}_{\text{STATE}}] / \_\_\_ \text{PP}_{\text{TIMEPERIOD}} \end{array}$

## 6.9 Conclusion

This article aimed to provide an insight in why and how West-Germanic posture verbs and motion verbs can get a bleached interpretation. We have seen that there are two phenomena at play. The first one is mere syntactic malleability, which we saw is at play for the posture verbs *to stand*, *to lie* and *to hang*. By changing the verbal superstructure one can turn an agentive, durative verb into a non-agentive, non-durative verbs. As such, the verb acquires the aspectual properties of a non-agentive state. Its aspectual properties then resemble the properties of copulas. As the locational prepositional phrase comes close semantically to a locational predicate, the copula-like interpretation should not come as a surprise. However, the occurrence of these verbs in small clauses indicates that they are not true copulas. They are merely lexical verbs put to use under a different functional structure. This possibility shows that agentivity and unboundedness, i.e. durativity, are not inherent properties of these posture verbs. They are properties of the functional structure with which they merge. We have called this phenomenon malleability and we have seen that these three verbs can undergo this kind of malleability not only in Dutch, but in all continental West-Germanic varieties under discussion, i.e. High German, Low German and Frisian.

The posture verb *to sit* and the motion verbs *to fall* and *to come*, in contrast, may grammaticalise into copulas. As a result, for these verbs at least two versions co-exist in the lexicon: for each of these verbs, there is a root and there is a copula. Whether such grammaticalisation has taken place depends both on the verb and the language. As such, we have found variation amongst the varieties under discussion.

Typologically, we must conclude that the continental West-Germanic languages are splitting languages. They have a default stative copula *to be*, but they may also have designated copulas for

specific semantic domains<sup>124</sup>. For locational predication, there is *to sit*, for temporal predication, there is *to fall* and to express origins, there is *to come*. Dutch has all three of them. Frisian, Low German and High German can also be characterised as splitting language, even though they have less copulas than Dutch. One might speculate that the other West-Germanic varieties might follow Dutch in the future.

---

<sup>124</sup> There are also non-stative copulas *worden* ‘to become’ and *blijven* ‘to stay’, but they are not relevant for the characterisation as a splitting language.

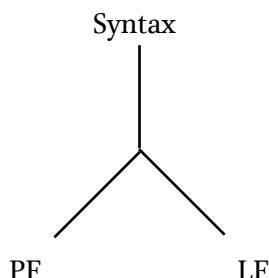
# 7 The Syntax-Semantics Interface in Distributed Morphology

## Abstract

Literature in Distributed Morphology tends to adhere to the principle of full compositionality at the LF-side: only a few authors have proposed post-syntactic structure-altering operations at the semantic side of the computation. Yet, the field does not generally assume a simple match between syntactic and semantic structure due to underspecification of the syntactic structure. The idea that syntax can be semantically underspecified has been explored in various ways. The idea goes back to the root node, which has been seen as underspecified since the seminal papers. This idea gradually spread to little v and, more generally, transitivising heads. However, if too many nodes in the structure are underspecified, the structure will be semantically underspecified. Nodes can only be underspecified if their context is sufficiently informative. This results in an unresolved issue in the field: we cannot assume that both the root and the transitivising node are underspecified.

## 7.1 Introduction

The Y-model entails three representations: there is the syntactic representation, the surface representation derived at PF and the semantic representation derived at LF. Given the Y-model, PF and LF representations do not interface:



Nevertheless, in an ideal scenario one expects them to match one another, given that they should both match the syntactic representation. The surface form should match the underlying syntactic structure because of the mirror principle (Baker 1985) and the interpretation at LF should match the underlying syntactic structure because of full compositionality.

Needless to say, this ideal scenario is commonly distorted. Indeed, it is well-known that the surface representation of a form does not necessarily match its meaning. Consider the following example. In plural DPs Standard German nouns will often be overtly marked for plural, as shown in (1a). However, in the presence of a diminutive affix, no plural marking is attested. As a result, the surface forms of the word *Kätzchen* in the plural DP in (1b) and in the singular DP in (1c) are identical.

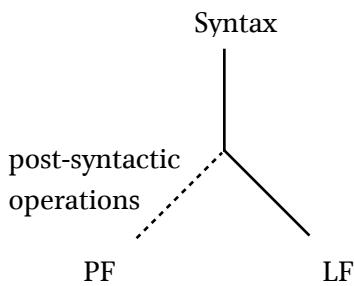
- (1) a. zwei Katz-en  
two cat-PLURAL  
'two cats'

- b. zwei Kätz-chen  
 two cat-DIMINUTIVE  
 'two kittens'

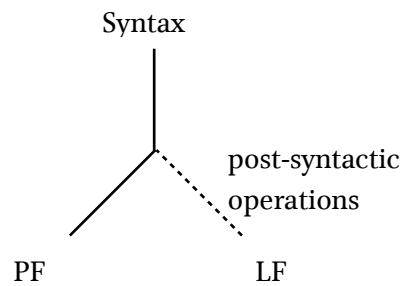
- c. ein Kätz-chen  
 one cat-DIMINUTIVE  
 'a/one kitten'

In short, (1b) does not show any overt plural marking, even though it clearly gets a plural interpretation. To account for such a mismatch, one could *in principle* formulate (at least) two hypotheses about the Y-model. Firstly, one could assume that a plural morpheme is present at the syntactic representation, but due to post-syntactic operations and rules at the PF-side of the Y-model it is not spelled out. This is illustrated in (2a). Secondly, one could postulate that a plural morpheme is not present in the underlying syntactic representation, but due to post-syntactic operations and rules at the LF-side of the Y-model, it does acquire a plural interpretation. This is illustrated in (2b). This situation would effectively constitute a syntax-semantics mismatch.

(2)a.



b.

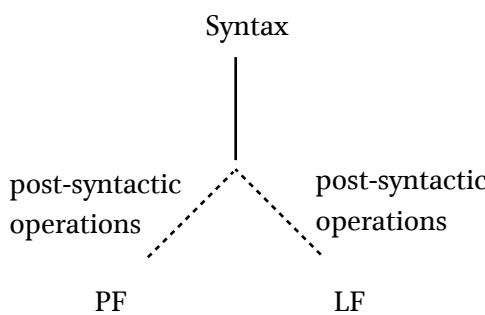


I believe that for the specific example in (1b), (nearly) all scholars working in Distributed Morphology would immediately accept a proposal along the lines sketched in (2a) and reject one as sketched in (2b) (and see De Belder 2011a for an analysis along these lines for the example in (1)). This assumption is based on two strong tendencies in the tradition of DM. Firstly, even though the mirror principle is a guiding principle in many DM analyses (see Harley 2012 for discussion), there is no denying that linguists in the DM framework have been most comfortable relaxing it. Since the seminal papers (Halle & Marantz 1993, 1994) there has been a long and steady tradition in the literature on Distributed Morphology discussing post-syntactic operations at the PF-side. As a result, there has been a general acceptance that post-syntactic morphological operations may alter the featural representation of the tree significantly, influencing how it is spelled out and thus allowing for considerable mismatches between surface structure and the syntactic representation. These operations are known as Fission (Noyer 1997), Impoverishment (Bonet 1991), Vocabulary Insertion allowing for the insertion of underspecified vocabulary items, and several Merge operations, including Morphological Merger (Marantz 1984), Syntactic Lowering (Bobaljik 1994) and Local Dislocation (Embick and Noyer 2001).

Secondly, one does not find structure-altering operations at the LF-side parallel to the ones proposed for the PF-side in early work on Distributed Morphology, apart from covert movement. DM literature tends to adhere to the principle of full compositionality at the LF-side, as discussed in section 2 of this article (and see Harley 2012 for discussion).

In sum, when a surface representation does not match the semantics of a structure, the go-to proposal would involve an operation at the PF-side of the Y-model. Nevertheless, *a priori* there is no principled reason not to explore the other logical possibility, as sketched in (2b). In fact, the opposition between (2a) and (2b) may be a false dilemma and the most natural assumption might be that operations at one side of the Y-model are mirrored at the other side, as illustrated in (3).

(3)



Indeed, mostly during the last decade, there have been empirical and theoretical explorations of mismatches at the other side of Y-model, i.e. between syntax and semantics. These proposals sometimes reflect ideas that have been formulated for the PF-side of the model, such as impoverishments. LF impoverishment (also known as ‘deprivation’), for example, entails that for some mismatches it is supposed that form and syntax are congruent and it is semantics that suffers from impoverishment. In such cases, deprivation deletes non-interpreted features that are supposed to be present at syntax and that are realized at the surface (Sauerland 2008, Nevins 2008 and see Heim 2008). The idea of LF impoverishment introduces the idea that just as there are post-syntactic morphological operations, there may be post-syntactic semantic operations. Other ideas include syntactic underspecifications, phase mismatches (Marušić 2005, Marušić & Žaucer 2008) and synonymous syntactic structures. However, in the last decade, mostly under the impulse of Alec Marantz and his students, a view has been developed that far surpasses the exploration of some ideas. It is proposed to sever semantics largely from syntactic features<sup>125</sup>. These ideas, the path that led to these ideas and the alternative view are presented in the current chapter.

This chapter is structured as follows. The following section presents the original view according to which meaning is derived compositionally from features present at syntax. Section 3 describes the view that syntax and semantics might function to some degree independently from one another and sketches the path that led to these proposals. It will become clear that the syntax of introducing the external argument lies at the core of the discussion. Section 4 very briefly addresses whether the operations Merge and Move are assumed to have semantic impact. Section 5 concludes.

---

<sup>125</sup> In the present contribution I use the words ‘feature’ and ‘head’ consistently in their original sense: heads are the structural positions in the tree that result from the operation Merge and features are the elements from UG that may be carried by those positions. However, given that one may posit that heads are created by merging features (from the numeration), the ‘head’ is understood as the syntactic position plus the features carried by that position. Furthermore, some authors will assume that a single head carries a single feature. In those contexts the concepts of heads and features are *practically* interchangeable. However, other authors might group several features on a single head. This certainly has been done in Distributed Morphology.

## 7.2 Full compositionality as the originally prevailing principle

As mentioned in the introduction, the original literature in Distributed Morphology tended to adhere to the principle of full compositionality at the LF-side. In this section I present three examples of trends in the literature that substantiate this claim.

A first example is that it is considered a quality if ‘a semantic theory is answerable to the morphology and if, conversely, a morphological theory is answerable to the semantics (Harbour 2008).<sup>126</sup> A correspondence between morphology and semantics is then established by proposing syntactic features that have a clear semantic definition and by decomposing syntactic properties in such features. Such analyses therefore rely on decompositions of syntactic concepts into features that are argued to be semantic primitives. *Person*, for example, is decomposed into features such as [+/-Participant in Speech Event] and [+/- Author of Speech Event] rather than assuming values such as 1, 2 and 3 (see, for example, Halle 1997). *Number* is decomposed into [singular] and [augmented], rather than assuming *singular*, *dual* and *plural* as primitives (Noyer 1997, Harbour 2008). The syntactic structures in these works are then built semantic feature per semantic feature in a strict compositional fashion. Morphological exponents are argued to realise (subsets of) precisely these features. By defining syntactic features with clear semantic content and by proposing a clear relation between these features and the exponents, a close match is proposed between the three representations, viz. the representations at Syntax, LF and PF. As such, these proposals aim to approach the ideal scenario sketched in the introduction. Many linguists working in DM would consider this a desirable result.

A second telling example of the tendency to focus on a compositional match between syntax and semantics is work on idioms in DM. It is emphasised that even though the meaning contributed by the Encyclopedia might be completely idiomatic, the meaning contributed by features at LF is invariably fully compositional. McGinnis (2002) shows that the aspectual properties of idioms are invariably fully compositional. The lexical properties of idioms may be irregular and stored, but their structural semantic properties, such as aspect, can be mapped onto their syntactic derivation in a fully compositional fashion. The same point is made in Marantz (1997), who states that ‘*transmission* can’t mean what *blick* could mean as nouns like *transmission* include an aspectual pre-verb, a verbal stem and a nominalising affix’ (Marantz 1997: section 1.3). Overriding the meaning of internal structure is simply not an option.

Thirdly, if two structures look identical at the surface but have a different meaning, the field usually attempts to show they have a different underlying syntactic structure, rather than postulating that syntactic structures can be ambiguous (see Harley 2012 for discussion and references therein). A lot of attention has been given, for example, to the structural difference between participles with eventive and stative interpretations (Marantz 1997, Embick 2004, Kratzer 2001, Alexiadou and Agnastopoulou 2008) and to various merge positions for the causative morpheme (Harley 2008 and references therein). These proposals go hand in hand with the more general view that the altering of the verb’s argument-structure is attributed to various functional heads in the syntactic domain, which operate in a strictly compositional fashion to derive inchoative, causative, applicative, active and passive variants of the verbal root (Pylkkänen 2008, see Harley 2012 for discussion). In the nominal domain, De Belder, Faust & Lampitelli (2014) discuss that cross-linguistically, one can structurally distinguish between diminutive and augmentative affixes that function as derivational affixes and those that merely realize a  $\text{Size}^\circ$  head.

---

<sup>126</sup> I am loosely citing Harbour (2008: 2-3) here.

Note that the tendency to value semantic compositionality so dearly is indeed a theoretical preference and by no means a logical necessity. In Nanosyntax, for example, form-based principles prevail more strongly. De Clercq & Vanden Wyngaerd (2017), for example, assume two projections in the comparative, which they call C<sub>1</sub> and C<sub>2</sub>, on the basis of arguments based on the surface representation, even though the presence of these projections is not motivated semantically. Form-based evidence suffices in Nanosyntax, even in the absence of semantic support. As such, the proposal contrasts sharply with, for example, Bobaljik (2012) in which the presence of every head is justified formally, syntactically and semantically. Even a proposal such as Embick (2010), which is morphophonologically in nature, only contains heads that are independently motivated at the syntax-semantics side.

In sum, semantic compositionality derived from syntactic features with clear semantic content is a strong guiding principle in DM. Nevertheless, within DM itself, this view has been challenged and alternative views have been developed. They are discussed in the next sections.

### 7.3 A semantically underspecified syntax

In the previous section, we have seen that overriding the semantics derived from the morphosyntactic structure is generally not considered an option in Distributed Morphology. What does seem to be an option, however, is a syntactic structure with an underspecified meaning. The idea that syntactic features have a clear semantic content certainly has been challenged in this way. In fact, the question whether the features of a head should *per se* determine the meaning of a head is a research question in its own right in DM. The mere idea that a syntactic node can be semantically underspecified was encoded in DM since its inception: the root node was conceived as an underspecified node. In more recent years, underspecification has been proposed for functional nodes as well. A central role in this debate is occupied by the functional head that introduces the external argument. This section sketches the debate and discusses the resulting proposals.

#### 7.3.1 The root as an underspecified node

In the seminal papers and in most work done in DM, the root is taken to be fully categorially and featurally underspecified. It has long been stipulated that it merely contains a feature [Root]. De Belder & Van Craenenbroeck (2015) take the idea of featural emptiness to the extreme and define the root node as an empty set as a consequence of the fact that the root is the first node in the tree. When starting the derivation of a syntactic structure, the operation Select will select a first functional feature  $\alpha$  from the Numeration. This feature is included into an empty derivation. This empty object is indeed empty, i.e. it is the empty set  $\emptyset$ . The very first Merge of the structure therefore creates the object  $\{\alpha, \emptyset\}$ . This empty set then functions as the root node. In their proposal the root is thus the most unspecified node imaginable. As such, they derive that roots have no grammatical features or a syntactic category, as argued for at length in Borer (2005a,b). Interestingly, they also define the root position structurally rather than lexically: a root is defined as a certain position in the tree. It is not defined by the lexical item that merges in the tree. They show that when functional items realize a root position, the functional properties of the lexical item are inert. After all, the properties of the root are defined by the absence of features in the tree, not by the vocabulary item that happens to realize it post-syntactically. The proposal also derives the fact

that roots are dominated by functional material (rather than the other way around). In sum, the structure guarantees that any root will appear as a blank sheet of paper and all syntactic properties of the structure will be defined by the functional superstructure.

Another possibility for the root node, commonly in use nowadays, is specifying the node with an abstract index that refers to a vocabulary item (Harley 2014), as in (4) (taken from example (15) in Harley 2014).

(4) PF instructions:

$$\sqrt{77} \leftrightarrow / \Theta_{\text{row}} /$$

LF instructions:

$$\begin{aligned} \sqrt{77} &\leftrightarrow \text{"vomit"} / [ v [ [ \_ ]_{\sqrt{ }} [ \text{up} ]_{\text{P}} ] ]_{\text{vP}} \\ &\leftrightarrow \text{"a light blanket"} / [ n [ [ \_ ]_{\sqrt{ }} ] \\ &\quad \{ \dots \text{other meanings in other contexts...} \} \\ &\leftrightarrow \text{"throw"} / \text{elsewhere} \end{aligned}$$

The index, e.g.  $\sqrt{77}$ , then merges at the root position. All phonological and semantic information to that node is only matched to the index at the post-syntactic interfaces. As shown in (4), these instructions -but not the root itself- may contain insertion contexts referring to categorial information. The index thus refers to phonological and semantic information which may contain categorial information, but the index itself does not have a single phonology, semantics, a category or other features. The root node remains underspecified. Notice how such a proposal manages to establish the integration of lexical knowledge in a non-lexicalist approach. The root node at syntax is not marked with any syntactic features and, as such, a lexical item does not determine the syntactic derivation. However, at Encyclopedia, we see a lexicon in full bloom. Lexical knowledge is thus certainly part of the linguistic system, but it comes too late in the derivation to affect the syntactic structure.

Looking back, we may note that the proposal by De Belder & Van Craenenbroeck (2015) is in principle compatible with Harley's (2014) view on Encyclopedia. Even though De Belder & Van Craenenbroeck (2015) does not assume the presence of an index at syntax, it does assume that a lexical item will come to realize the root node at Vocabulary Insertion. The syntactic structure enriched with lexical items may then interface with Encyclopedia. More generally, one does not necessarily need to assume the presence of a root index at syntax to include a rich view on lexical knowledge in the linguistic system. In other words, there is no contradiction between a radical anti-lexicalist view on syntax and acknowledging encyclopedic and lexical knowledge.

In short, the question whether the features of a head should determine the meaning of a head has thus been responded to with a resounding no for the root. As such, the root node is important to understand the possibility of underspecified nodes in DM. Since the early days of DM the way the root has been given a place in syntactic structures illustrates what an underspecified node depends on: if the surrounding structure is sufficiently specific, a free riding node does not undermine the interpretability and well-formedness of the syntactic structure at the interfaces. The emptiness of the root node is possible precisely because the functional structure is such a rich composition of features. For example, if the functional structure of a verb is fully equipped to derive a causative verb, there is no need to encode causativity on the lexical node. In fact, there is even no need to

encode verb-hood on the lexical node. The avoidance of encoding information more than once in the syntactic tree is a typical tenet of work in DM.

### 7.3.2 Little v

In Distributed Morphology, the external argument is not introduced by the verb but by a separate functional head. After all, the ‘verb’ is not a primitive notion in DM: the verb is a root merged under a verbal functional spine. One of the heads in this functional sequence will thus introduce the external argument in its specifier. There is some disagreement on which head exactly fulfils this role. Some authors point at little  $v^0$ , i.e. the head that verbalises the structure, others point at Voice, i.e. the head that determines, amongst other things, whether the structure is active or passive. I will not go into this issue any further. However, the following paragraphs discuss both little v and Voice interchangeably, depending on whose work is under discussion. The reader should keep in mind that what unites both heads is that they are supposed to introduce the external argument and this section is in essence about the nature of functional heads that introduce arguments.

Work on the head that introduces the external argument raised important issues on the relation between syntax and semantics. The original, empirical core of the problem is that the external argument can adopt various semantic roles. These roles may depend on the verb class and the argument structure of the sentence. Verb classes determine possible argument structures and their semantic roles (Levin & Rappoport Hovav 1995). For example, *wither* implies an internal cause, but a verb such as *destroy* requires an agentive, external cause of the event and is incompatible with an internal cause (Levin & Rappaport Hovav 1995; Reinhart 2002; Alexiadou, Anagnostopoulou, and Schäfer 2015).

- (5) The flowers wither.
- (6) The soldiers destroyed the city.
- (7) \*The city destroys.

Similarly, semantic roles may alter in various syntactic structures. In the causative alternation, for example, a single verb may occur both in a transitive and an intransitive sentence (see Levin 1993; Schäfer 2008; Alexiadou, Anagnostopoulou, and Schäfer 2015 and references therein, example taken from Alexiadou, Anagnostopoulou, and Schäfer 2015:2).

- (8) Mary opened the door.
- (9) The door opened.

Crucially, in the transitive, causative sentence, the external argument will be understood as the cause of the event. In the intransitive, anti-causative sentence, the external argument is understood as a theme that undergoes the event.

Now, if various kinds of external arguments exist, what does that imply for the functional head that introduces the external argument? Does it determine the semantics of the external argument? If so, how can a single head determine various meanings? The problem and some seminal ideas are introduced in Marantz (1997). He distinguishes the agentive reading of *grow* from the internally caused reading syntactically by means of two different little v’s: an agentive one and a non-agentive one.

- (10) \*John grows tomatoes. [agentive reading, agentive v]
- (11) The tomatoes grow. [internal cause reading, non-agentive v]

He proposes that these two v's could either be understood as flavors of one another, i.e. as subtypes of little v, or that a single head could optionally project an agent. This second proposal should be understood as an underspecified little v: a part of the semantics of a functional head is but optional and thus underspecified.

The difference between the two ideas might seem subtle, but it does reflect a very different view on functional heads. Under the 'flavors approach', the inventory of functional heads is conceived as a rich array. Indeed, every semantic difference may be encoded by a different functional head. The word 'flavors' is in this respect not a bad metaphor: it does trigger this image of an Italian ice cream shop with hundreds of choices in flavors to compose a unique cone. As such, the spirit of the flavors approach contrast sharply with the 'underspecification approach'. Under this approach, the inventory of functional heads can be conceived of as absolutely minimal, thanks to the underspecification: indeed, there is only one lean, versatile functional head to introduce arguments (Wood & Marantz 2017:2).

We will see in the next paragraphs that the idea of adopting flavors has been explored at the turn of the millennium (Folli & Harley 2002), to be abandoned later on. The idea of an underspecified head, in contrast, won gradually more ground and has been extended beyond little v into the more general idea that syntax is autonomous from semantics (Wood 2015, Myler 2016, Wood and Marantz 2017). The following sections describe this evolution, the various ideas that have been presented and the current outcome.

### 7.3.3 Flavors

Folli & Harley (2002) proposed that little  $v^0$ , i.e. the feature on the head that introduces the external argument according to these authors, may come in several subtypes. More specifically, they proposed that this feature exists in two subtypes, viz.  $V_{do}$  and  $V_{cause}$ <sup>127</sup>. It is thus not the case that little v in their proposal may combine with various other features to form semantically different heads. Little v itself was taken to exist in several subtypes. They referred to these subtypes as 'flavors'. The various flavors of this feature echoed various theta-roles of the external argument:  $V_{do}$  introduced an animate Agent role, i.e. a role to which intentions can be ascribed, whereas  $V_{cause}$  introduced any external argument that could be understood to cause the event expressed by the verb. It was conjectured that specific verbs could select for specific flavors of little v. For example, consumption verbs could only select  $V_{do}$  and not  $V_{cause}$  in a non-resultative context. As such, the proposal captured the fact that only animate, agentive subjects are licit external arguments for such verbs:

- (12) Mary ate the ice cream.
- (13) \*The sea ate the beach.

The approach was abandoned in later work. Harley (2013, 2017) presents further data which led her to conclude that causation is introduced by its own dedicated head and should not be considered a subtype of the external argument introducing head (and see Pylkkanen's 2008). For example, Harley (2017:11) presents the following example from Hiaki in which it is shown that the exponent *ta*, realizing the causative morpheme, can co-occur with the exponent *wa*, which realizes

---

<sup>127</sup> To be precise, they also proposed  $V_{become}$ , but it was less central in the discussion.

the passive morpheme in Hiaki. This indicates that the causative head and the Voice head are separate heads:

- (14) U vaaso ham-ta-wa-k.  
The.NOM glass break-TR-PASS-PRF  
'The glass was broken.'

Similarly, *ta* can co-occur with an exponent *tua*, which indicates that the causation is direct. Direct causation implies that an agent, here marked as an object, is introduced into the structure. Again, causation and agentivity are realised by separate morphemes<sup>128</sup>:

- (15) Juan Maria-ta vaso-ta ham-ta-tua-k.  
Juan Maria.ACC glass.ACC break-TR-CAUS.DIR-PRF  
'Juan made Maria break the glass.'

Crucially, these examples show that causation and agentivity are marked by separate morphemes, indicating that they constitute separate heads in the syntactic structure.

Yet, for some languages it may seem that a single head is responsible both for causation and the introduction of the external argument. Harley (2017) therefore argues that when several features seem to be realised by a single head, it is because several heads are collapsed into a single head due to a bundling parameter. More specifically, she adopts Pylkkänen's (2008) proposal that a parameter exists that determines whether the head that introduces the external argument, checks accusative case and finishes the cycle and the head that verbalises the root collapse into a single head in a given language. This bundling idea goes back to the Split-IP parameter proposed by Thráinsson (1996) and Bobaljik and Thráinsson (1998). Such parameters thus have a long and apparent steady history in DM. For an example of such a bundling parameter in the nominal domain in DM, see De Belder (2011b).

### 7.3.4 Expletive Voice

Extensive work on causation has been done by Alexiadou, Anagnostopoulou and Schäfer (2015)(henceforth AAS 2015), building on work by the same authors (Schäfer 2008 a.o.). Central in their work is the contrast (see also above) that exists between externally caused events, called causatives, as in (16) and internally caused events (17), called anti-causatives.

- (16) Mary opened the door.  
(17) The door opened.

The main idea is that, in a language such as English, a specific head Voice sets these two structures apart. Voice is present in causatives. It introduces an external argument which is understood as the agent that causes the event expressed by the verb and its internal argument. Voice is absent in anti-causatives. As no external cause is introduced in the argument structure of the verb, the event the theme undergoes is understood to be caused internally. AAS (2015: chapter 2) demonstrate

---

<sup>128</sup> See Harley (2017:11) for a contrasting example of indirect causation.

extensively, that, indeed, anti-causatives test negatively for all evidence that an external argument is explicitly or implicitly present in the structure.

For English, the scenario is thus clean and simple when it comes to the syntax-semantics interface: the presence or absence of a head results in a semantic difference in argument structure that straightforwardly follows from the nature of the head. However, there is a considerable complication once other languages are taken into account and the solution proposed by AAS has a considerable impact on the relation between syntax and semantics, as detailed in this section.

Given that anti-causatives are argued to lack Voice, the prediction follows that they will always lack any Voice morphology. However, this prediction is not borne out. AAS (2015) show that anti-causatives in various languages may show morphology that is independently argued to reflect Voice: in Romance languages and in German they show reflexive clitics or pronouns, in Greek they show non-active morphology. The morphology thus does suggest Voice is present in these anti-causatives: the morphemes may be syncretic with passive morphology, as in Greek, or the sentences may show transitive structures, such as the reflexive structures in German. Yet, there are consistently no truth-conditional meaning differences between marked and unmarked anti-causatives. The morphosyntax of such anti-causatives thus does indicate the presence of Voice, but their semantics contradicts it.

To resolve this issue, two solutions are immediately rejected by AAS. Firstly, one could deny that the attested exponents realise syntactic structure. This is rejected because of the realisational nature of DM: morphemes are assumed to realise structure. Secondly, one could argue that the problematic piece of structure actually detransitivises the structure. This solution is rejected because of the Monotonicity Hypothesis (AAS 2015 refer to Kiparsky 1982, Doron 2003, Koontz-Garboden 2009, Schäfer 2013 and Kratzer 2013): structure-building cannot take away meaning. Indeed, we have seen above that DM will never assume that structurally built meaning can be overridden.

AAS thus bite the bullet and accept that marked anti-causatives may contain a syntactic layer that is absent in unmarked anti-causatives, albeit one that is thematically and semantically inert. They thus argue that Voice can be inert. Such a Voice projection is called Expletive Voice.

Technically, differences between various types of Voice are regulated by means of features. For example, a single Voice head in Greek will be marked with different features when being part of a passive sentence than when it is part of an anti-causative. Note that this is thus certainly not a flavor-type approach. There is but a single Voice head. However, it may be marked with optional features, such as ‘agent’ or ‘causer’. In the absence of any features, the Voice head will be an expletive Voice (AAS 2015:101).

The proposal has two important theoretical consequences. Firstly, there are two ways to derive an anti-causative: in an unmarked anti-causative the Voice head is simply absent. In a marked anti-causative a semantically inert, expletive Voice head is present. This state of affairs is accepted by AAS (2015): different syntactic structures may end up being synonymous. It is worth noting that this proposal dovetails in this respect with the more general acceptance in DM that various syntactic structures can have the same meaning, both across languages and within a single language (Embick & Marantz 2008, Kastner & Zu 2017): DM allows various synonymous syntactic structures.

Secondly, the semantic and syntactic properties of the Voice head are teased apart (see also Schäfer 2008). Voice is the locus of both semantic and syntactic transitivity, but these notions do not need to coincide. They do coincide for active causative verbs, which are both syntactically and semantically transitive. The syntactic transitivity is then achieved through the presence of Voice and the semantic transitivity through thematic features. The notions also coincide for unmarked

anti-causatives, which are both syntactically and semantically intransitive, as they lack both Voice and any thematic features. Marked anticausatives, however, are syntactically transitive, but semantically intransitive. The morphological realisation of Voice then depends on syntactic transitivity: it reflects the presence of the Voice head and, for some languages, Voice's specifier. It does not necessarily reflect transitivity in the semantic sense of the word.

AAS (2015) thus propose that a syntactic head may be semantically inert. Note that they did not give up on the general idea that meaning is built compositionally by means of features. If the structure does contain an external argument with clear thematic semantics, the semantics is still derived compositionally through features. However, crucially, a head is not necessarily defined by its semantic features and as such, they may be absent: in the absence of any semantic features, the Voice head is still defined, its essence being a syntactic transitiviser. (The head is thus probably only marked with its own defining syntactic feature which has no semantic value.) The great shift one witnesses in the work by AAS (2015) is thus certainly not that meaning is no longer derived featurally, but rather the fact that syntactic and semantic information can be severed. As such, syntactic structure that does not carry meaning becomes conceivable and Voice can be defined as an underspecified head.

### 7.3.5 Contextual allomorphy

Wood (2015) presents an approach to Voice very much in the spirit of AAS (2015). He adopts the idea that semantic information is severed from syntactic structure-building and an underspecified Expletive Voice is a central concept in his work as well. The major difference with AAS (2015) is that he pushes the proposal yet one step further. Whereas AAS (2015) still ascribed to the idea that meaning is compositionally derived from features, Wood proposes that meaning is derived from the context. It is at this point that the approach drifts away completely from the original prevailing idea presented in section 2 that meaning is built fully compositionally by means of features in the syntax.

Recall that AAS (2015) proposed that Voice may or may not be present in anti-causatives. When present, it is semantically inert and therefore called Expletive Voice. Wood, in contrast, argues that Voice is invariably present in anti-causatives. In unmarked anti-causatives he assumes an Expletive Voice which does not project a specifier, for marked anti-causatives he postulates a Voice projection that does project a specifier, by virtue of a selectional feature requiring a DP. However, the D that merges in the specifier will be an expletive. He proposes that, for example, the *-st* clitic in Icelandic fulfills this expletive function:

- (18) Hurdin opna-di-st.  
door.the.NOM open-PAST-ST  
'The door opened.'

Wood focuses on Icelandic anti-causative constructions. This language has both marked and unmarked constructions. Marked constructions can either be expressed through Voice morphology or by a filled Spec, Voice, as in the example above. The details should not concern us here. However, what is relevant here is how it is determined in Icelandic whether a verb will occur in an anti-causative construction at all and if so, in which kind of anti-causative construction it occurs. A related question is how Voice will be interpreted. After all, if there are no semantic features on this head to hint whether the external argument should be interpreted as fulfilling a thematic role and, if so, which one, where does the correct semantic interpretation come from?

Wood answers both questions in a similar fashion: both the distribution and interpretation of the root and the interpretation of Voice depend on the context. The relevant context is the entire vP. Let us see how he proceeds. Firstly, he adopts Harley's (2014) proposal that roots are but an index at syntax which point to both PF instructions and semantic instructions. The example of the hypothetical root  $\sqrt{77}$  given above is repeated here. Recall that it shows that an index will be linked to instructions at the interfaces in Harley (2014):

- (19) PF instructions:

$$\sqrt{77} \leftrightarrow /θrow/$$

LF instructions:

$$\begin{aligned} \sqrt{77} &\leftrightarrow \text{"vomit"} / [ v [ [ \_ ]_{\sqrt{}} [ up ]_P ] ]_{vP} \\ &\leftrightarrow \text{"a light blanket"} / [ n [ [ \_ ]_{\sqrt{}} ] \\ &\quad \{ \dots \text{other meanings in other contexts...} \} \\ &\leftrightarrow \text{"throw"} / \text{elsewhere} \end{aligned}$$

Applied to the Icelandic root  $\sqrt{\text{blóm}}$ , for example, Wood proposes the following instructions:

- (20) PF instructions:

$$\sqrt{42} \leftrightarrow /plom/$$

LF instructions:

$$\begin{aligned} \sqrt{77} &\leftrightarrow \text{"bloom (literally)"} / [ \text{Voice}_{\{D\}} [ [ \_ ]_{\sqrt{}} v ] ] \\ &\leftrightarrow \text{"bloom (metaphorically)"} / [ \text{Voice}_{\{D\}} [ [ \_ ]_{\sqrt{}} v ] ] \\ &\leftrightarrow \text{"flower"} / [ [ \_ ]_{\sqrt{}} n ] \\ &\quad \{ \dots \text{other meanings in other contexts...} \} \\ &\quad \text{No elsewhere interpretation} \end{aligned}$$

The LF instructions state that the Icelandic root  $\sqrt{\text{blóm}}$  has a designated interpretation under  $\text{Voice}_{\{D\}}$ . There is also a nominal interpretation, but no elsewhere interpretation. One could, of course, merge  $\text{Voice}_{\{D\}}$  on top of a vP containing the root  $\sqrt{\text{blóm}}$ , but the consequence will be that the construction is uninterpretable given the lack of a default semantics. As such, Wood derives the fact that this specific root is tied to a specific anti-causative construction. He explicitly states that he intentionally recasts selectional features as LF instructions. It is crucial that the entire vP is part of the LF instructions: it is a central idea in Wood that the syntactic structure determines the interpretation.

We now arrive at the part of his proposal which is crucial for the discussion at hand: how is the relevant interpretation for Voice reached in this proposal? In other words, how is the understood thematic role assigned? Consider the following observation: just as internally caused events in English and other languages, a blooming event will be incompatible with an external causer in Icelandic. The anti-causative will be marked with the expletive clitic *-st*, which as we know by now, occupies  $\text{Spec}, \text{Voice}_{\{D\}}$  and thus implies the presence of  $\text{Voice}_{\{D\}}$ .

- (21) \* Sólin hefur blómgað seljuna.  
sun.the.NOM has bloomed sallow.the.ACC

- (22) Seljan                  hefur blómast.  
           sallow.the.NOM has bloomed-ST  
       ‘The sallow has bloomed.’

The interpretation of the vP goes as follows, according to Wood. Upon the merger of  $\text{Voice}_{\{\text{D}\}}$ , vP is spelled out. Because of the presence of v, the structure is interpreted as a change-of-state event, following a proposal in Marantz (2009, 2013) in which it is argued that a single v head with a DP as its complement is the structure of a change-of-state event. The LF instructions dictate that the root in this verbal structure is to be interpreted as a blooming event. Given that the event applies to a sallow, the literal allosememe will be selected. This allosememe refers to an internally caused event. This is simply implied by the fact that “blooming” describes an internally caused event (Levin & Rappaport Hovav 1995). Hence,  $\text{Voice}_{\{\text{D}\}}$  is interpreted as expletive Voice, leading to a successful interpretation of an anti-causative in (22) and to a crash in (21). The thematic role to be assigned to Voice, if any, is thus computed post-syntactically, not on the basis of syntactic features, but on the basis of syntactic context and root allosemes. Meaning is thereby certainly not structure-independent: the LF instructions clearly refer to structure. However, the meaning is not *built* by the structure or by features. As such, the proposal clearly severs semantics from syntax.

Note that the proposal is in a way intrinsically comfortable with the fact that the syntactic structure even contra-indicates the interpretation that is eventually achieved. The original intuition behind Voice was that a certain head could transitivize the structure, thereby introducing an agent. Now, under this proposal, the presence of Voice is not inherently tied to an agent at all: whether a structure implies an agentive role is but to be established through an LF instruction. Note in this respect, that AAS (2015) will only assume the presence of Voice in anti-causatives if there are morphosyntactic indications to do so, whereas Wood simply assumes its presence across the board. The work by AAS (2015) thereby still reflects the idea that merging Voice is *in se* odd in anti-causatives and should not be assumed beyond necessity. This hesitation to assume the merger of a head the nature of which is at odds with the meaning is given up in Wood’s proposals, stressing the autonomy of syntax.

### 7.3.6 The semantically underspecified head $i^*$

We have seen that AAS (2015) proposed to treat Voice in itself as a semantically inert, underspecified head of which the *raison d'être* is merely to transitivize the structure. We further witnessed how Wood proposes to replace semantics derived from syntactic features by semantics derived from the syntactic and lexical contexts. Once one thinks of these possibilities, there is no reason not to take it one step further. If an underspecified head can introduce an external argument and if its meaning thereby relies on its context, why would it not be able to introduce other arguments as well? After all, the context indicates how the participant is to be related to the event.

Wood and Marantz (2017) present such a proposal. In Distributed Morphology, various heads have been taken to introduce arguments. There is of course the Voice head, but there are also applicative heads, prepositions which introduce non-core arguments and little p, which introduces the external argument of prepositional phrases. Wood and Marantz (2017) argue that all these heads can be reduced to a single head: there is but a single, semantically underspecified head that adds participants to events. They call this head  $i^*$ . Little p is replaced by  $i^*$  merged with a pP, Voice is redefined as  $i^*$  merged with a vP and the Low Applicative is replaced by  $i^*$  merged with a DP.

Prepositions and High Applicatives are understood as  $i^*$ 's that merged with roots, an option to which I come back below.

The head  $i^*$  is categorially underspecified and merely serves to select a DP. The head does not remain categoriless throughout the derivation: it will inherit its category from the first constituent it merges with. Its distribution thus does define its category. Technically, this is understood as feature valuation:  $i^*$  comes with an unvalued categorial feature, which is then valued by its sister.  $i^*$  can merge with all kinds of categories, so it is categorially quite chameleonic.  $i^*$  derives its meaning in all its categorial appearances through contextual allosemy, as described in the previous section. For example, in the context of a locative PP it may be interpreted as a head that introduces a figure, in the context of an agentive vP it may be interpreted as a head introducing an agent. At LF the relevant interpretation rules are context-sensitive:

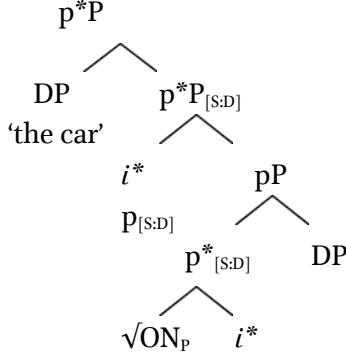
- (23)  $[[i^*]] \leftrightarrow \lambda x \lambda s. \text{FIGURE}(x, s) / \_\_ \text{ (locative PP)}$
- $[[i^*]] \leftrightarrow \lambda x \lambda e. \text{AGENT}(x, e) / \_\_ \text{ (agentive vP)}$

The interpretation rules in (23) only illustrate a fragment of all the possible interpretations of  $i^*$ . For each category  $i^*$  may have a range of meanings, such as various thematic roles, including an expletive meaning, see the previous section for an example and details. Note that such rules are an innovation. In DM approaches in which meaning is built compositionally on the basis of features in syntax, the exact meaning of a functional head will not need to be stipulated as a context-sensitive rule at LF as there is nothing context-sensitive about a compositionally built meaning and as the meaning of features is given by UG. It is unclear to me whether the context-sensitive rules presented by Wood & Marantz are meant to be universal, with perhaps some parameters, or learned. In older proposals, in which semantics was derived straightforwardly from syntactic features, this question did not arise: it was self-evident that the features were defined by Universal Grammar.

Notationally, the head  $i^*$  is described as  $\{\text{[CAT: } \_\_, \text{ [S:D]}\}$ . The notation  $[\text{CAT: } \_\_]$  should be read as 'a head with an unvalued categorial feature',  $[\text{S:D}]$  should be read as 'the selectional property ("S") of this head is such that it selects for a constituent of the category D. The asterisk on  $i^*$  is meant to describe that the head will close off the extended projection of the first constituent it merges with.

Interestingly,  $i^*$  may merge with lexical roots to derive prepositions and High Applicatives. As such, even lexical information may affect the interpretation of  $i^*$ . Wood and Marantz define a preposition as an  $i^*$  head merged with a prepositional root of category P. Consider how they derive the constituent *the car on the road*. A prepositional root adjoins to  $i^*$ . The root's categorial feature values the unvalued categorial feature of  $i^*$ :  $i^*$  becomes a preposition. A prepositional root and  $i^*$  together thus combine to form a preposition. The root defines the lexical semantics of the whole: the meaning of *on* defines *the road* as a surface and the prepositional phrase as a locative PP.  $i^*$  contributes its syntactic property of argument-selection. As such, it may select *the road*. A second  $i^*$  is merged. Thanks to the LF instruction stipulating that an  $i^*$  which merges with a locative PP is to be interpreted as a figure, *the car* is effectively interpreted as such.

(24)



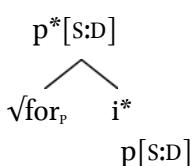
A High Applicative is defined as an  $i^*$  head merged with a category-neutral prepositional root. For the sentence *John held the purse for Mary*, in which *for Mary* can be interpreted both as a beneficiary  $pP$  and a beneficiary high applicative, Wood and Marantz thus, in effect, propose that to derive the prepositional reading  $i^*$  merges with a prepositional root 'for' of category  $P$  and to derive the applicative reading it merges with a category-neutral prepositional root 'for'.

Recall from section 3.1 that the root node is important to understand the possibility of underspecified nodes in DM as it illustrates that the possibility of underspecified nodes has been encoded in the theory from the start. To wit, the proposal in Wood & Marantz 2017 to adopt a single argument-introducing head  $i^*$  is reminiscent of the way the root node has been conceived. Both  $i^*$  and the root node are categorially and semantically underspecified and depend on the syntactic context to be interpreted at the semantical interfaces. The root node therefore has been conceived as a featureless, acategorial node and similar arguments of economy support  $i^*$ .

Ironically, the price for the underspecification of  $i^*$  is paid by the root. Firstly, we have seen that Wood (2015) derives the exact interpretation of Voice, the precursor of  $i^*$  in his work, via the meaning of the root. For example, if the root contains an LF interpretation that implies an internally caused event, Voice will get the semantically inert expletive interpretation. This proposal clearly moves away from root-dominant proposals in which roots are conceived of as fully malleable under a functional sequence that uniformly and compositionally determines their meaning (see, for example, Borer 2005a,b, 2013a).

Secondly, Wood & Marantz (2017) propose that a root may be marked with a categorial feature. For example, when  $i^*$  needs to be interpreted as a little  $p$  in the structure, it needs to merge with a head that can mark its empty categorial value as prepositional. However, in this case  $i^*$  merges directly with the root. The solution is to assume 'prepositional roots', i.e. roots that are marked with a prepositional value (Wood & Marantz 2017, fragment of their example (7)):

(25)



Consider what is going on here. There is a lexical core *for*, the functional information that the syntactic chunk in which it is embedded will need to function as a preposition and the syntactic information that *for* selects a complement. There are two roads to follow in DM: one can assume a root and a little  $p^\circ$ . The little  $p^\circ$  is responsible for the categorial information and the complement-selection. The root is inert and empty. Wood and Marantz (2017) divide the labor otherwise: the root

carries the categorial information and  $i^*$  merely selects the complement. All DM approaches will share the insight that it would be superfluous to assign the categorial feature to more than one node and that some node thus gets the privilege of being underspecified. However, at least one node will have to carry the categorial feature. Assuming that the category-assigning head is categorially underspecified will thus result in a more specified root node.

Thirdly, recall that Wood and Marantz assume the existence of a prepositional root ‘for’ of category P for the prepositional reading and of a category-neutral prepositional root ‘for’ for the applicative reading. This goes against the spirit of roots: the concept of roots reacted against multiplying lexemes, putting the burden of categorial ambiguities in the functional sequence. Here the opposite happens: the functional sequence is so underspecified, the root becomes specified and multiplied again.

### 7.3.7 Conclusion

We have seen in this section that the various thematic roles associated with the external argument gave rise to an extensive empirical and theoretical debate in Distributed Morphology. It is a valid question whether syntax should determine the semantics of this argument. Three approaches have been proposed to tackle the issue. Firstly, one has argued that functional heads come in various subtypes, each subtype being marked for its own semantic flavor. This approach did not survive the test of time. Secondly, there is the trusted approach to build the various meanings in syntax compositionally by means of features. This approach is deeply embedded in the tradition of DM as presented in section 2. Harley (2017) is an example of this approach: each semantic component is introduced by a feature on a designated functional head. Various heads may collapse, resulting in various features on a single head. Thirdly, there is the innovative approach to build syntactic structures with fully underspecified functional heads. Meaning is then derived from the context. The  $i^*$  head in Wood and Marantz (2017) is a clear example of extreme underspecification. The approach presented in AAS (2015) is a hybrid solution which, as a precursor of the third approach, combines elements from the second and the third approach. Meaning is still assumed to be derived from syntactic features, but heads are severed from their features. As such, featureless, and hence, meaningless heads became an option.

It is at this point hard to predict which approach will eventually dominate the field. However, I do believe the choice might depend on the theoretical understanding of the root in the future. As I have pointed out, underspecified functional heads put a heavier burden on the root: the functional head depends on its context to derive meaning and one element in the context might be the root. As such, the meaning of the root will have to feed the semantics of the functional structure. This approach is not only at odds with many proposals embedded in Distributed Morphology, but also with root-based work outside of Distributed Morphology, such as Borer (2005a,b, 2013a), which emphasises that it is the root’s meaning which is malleable under a semantically strictly defined functional structure. It is an open question whether it will prove to be more adequate to underspecify the root or the argument-selecting functional heads. Note that the discussion does not only entail consequences for the syntax-semantics interface, but also for the lexicon-syntax interface. Putting a burden on the root to define the properties of the syntactic structure is by definition the more lexicalist approach.

#### 7.4 Move and merge with(out) meaning

For reasons of space, this chapter only dealt with the question whether meaning is derived from syntactic features. We failed to discuss the semantic effects of the syntactic operations Merge and Move. One may wonder, firstly, whether phrases are interpreted where they merge and/or move to and secondly, whether head movement has semantic implications.

When it comes to phrases, I would like to mention that the saturation of thematic roles can be subject to what is called ‘delayed gratification’. Wood (2015) and Myler (2016) argue that specifiers can be arguments of a certain head semantically, without being the syntactic specifier of that head. These proposals are cognate to Ramchand (2008), in which it is argued that specifiers are allowed to move from the specifier of one functional head to the specifier of another one, hence serving as the argument of more than one predicate. I refer the reader to those sources for more information.

When it comes to heads, I would like to point out that the field has no definitive answer. Head incorporation is often seen as having no semantic effect at all: there are no differences between words and non-words (Bobaljik 2012). In other words, two different syntactic structures (with or without movement) may be perfectly synonymous. This result is seen as a desirable effect (Harley 2012). Nevertheless, some authors will exploit head movement as a semantic tool. Bruening (2010:551), for example, proposes that in chains built by head movement, the heads that are involved can be interpreted in any of the positions they move through.<sup>129</sup>

There is further a complication with this dominant view when it comes to compounds. Outside of the domain of Distributed Morphology head incorporation is usually employed precisely to derive the syntax *and* semantics of compounds (Mithun 1984 and subsequent literature). Semantically, incorporation then derives two semantic properties for compounds: the non-referentiality of the non-head and the word-worthiness of the compound. The word-worthiness is a concept related to interpretation. It refers to the fact that the compound will only be felicitous if it refers to an institutionalized or socially salient activity or object. In DM, incorporation solely derives the syntax of compounds (Harley 2009). The non-referentiality of the non-head will be ascribed to the absence of a D-layer and the word-worthiness of the compound needs to be stipulated at the encyclopedia. However, note that the incorporation is not completely inert. It will derive a word syntactically, which will be recognized as such by Encyclopedia. In short, Distributed Morphology has no definitive unified view on what incorporation may establish semantically.

#### 7.5 Conclusion

This chapter has presented two ways to approach the relation between syntax and semantics in Distributed Morphology. The first approach has the longest tradition in Distributed Morphology. It assumes that surface morphology reflects the presence of heads and features. These innate features are assumed to have semantic implications and they build the meaning of the sentence compositionally. If there is a clear mismatch between the surface representation and the meaning, it is usually assumed that the distortion took place at the PF-side of the tree. The second approach is a younger alternative. It severs semantics from syntax. It entails that the presence of a syntactic head does not necessarily have a semantic impact: heads may be expletive heads. In such cases, the syntactic head is underspecified. The possibility of underspecified heads led Wood & Marantz

---

<sup>129</sup> This idea is reminiscent of Ramchand’s (2008) specifier movement.

(2017) to propose a completely underspecified head to introduce participants to events. This featurally and categorially underspecified functional head is reminiscent of the way the root node has been conceived in Distributed Morphology. Yet, it is important to note that a syntactic structure can probably not afford to have nothing but underspecified nodes. As such, adopting underspecified functional nodes goes hand in hand with assuming richer root nodes.

To date, the alternative view co-exists with the original view and has not replaced the original view. For example, Wood & Marantz's (2017) view on underspecified prepositions, which, as we have seen, are reduced to  $i^*$ , exists alongside more old-style semantic compositions of p, such as Levinson 2011. Note that a more intermediate position, as presented in Alexiadou, Anagnostopoulou & Schäfer (2015), is easily reconcilable with the original view. After all, it still adheres to the view that meaning is basically derived from semantic features. The view that, in the absence of such features, there is an absence of meaning -hence deriving an expletive syntactic head-, actually simply follows from the original assumptions. In contrast, the emphasis on an independent semantics, which relies on the syntactic and lexical context rather than on features, as emphasised by Wood, is less easily reconcilable and thus forces us to establish which alternative is indeed the most adequate one.

## **8 A Split Approach to the Selection of Allomorphs**

### **Vowel length alternating allomorphy in Dutch**

#### **Abstract**

In this article it is argued that the selection of allomorphs is distributed over two modules, viz. Vocabulary Insertion and Phonology. This is done on the basis of a case study of vowel length alternating allomorphs in Dutch. The data show a split pattern: some empirical domains can be fully captured by phonological principles. For these cases, the phonologically most optimal allomorph will be selected. In other empirical domains, phonological principles still account for many of the attested data. Yet, one attests lexicalised exceptions as well, which are clearly phonologically non-optimal. The data echo opposing views in the literature: some proposals attempt to reduce allomorph selection to phonology, others focus on the fact that many examples are simply not phonologically optimal and suggest that allomorph selection should not be done by Phonology. I argue that the opposing nature of these two types of data is actually indicative of the way the selection of allomorphs is organised. More specifically, both Vocabulary Insertion and Phonology can determine the selection of allomorphs. Vocabulary Insertion is responsible for stored information, Phonology is responsible for phonologically optimising patterns.

#### **8.1 Introduction**

Stem allomorphy may occur when for a single root two (or more) variants have been stored. These variants are thus available. Yet, the mere fact that an allomorph is available does not necessarily imply it will be put to use. In order for the allomorph to surface, it should be selected. This article is an investigation into the selection mechanisms that determine the insertion of stem allomorphs, the order in which these conditions operate and their place in the Y-model.

Hypothetically, various principles may select an allomorph. Firstly, the allomorph may be selected to satisfy phonological requirements (Paster 2006; Anderson 2008; Kager 2008; Wolf 2008; Nevins 2010). Secondly, lexicalised contexts may determine the use of an allomorph, as its use may be associated with a specific idiom, collocation or other lexical context. Finally, syntactic contexts may select an allomorph; it may spell out specific syntactic features (cf. Lowenstamm 2007; Acquaviva 2008; Hermans and Postma 2009; Kramer 2009; Bye and Svenonius 2010) or the allomorph may simply be associated with a specific syntactic context, without spelling out a specific head (Harley 2014).

In this article I discuss an instance of allomorphy in Dutch. I show that both lexicalised insertion conditions and phonological conditions may determine the selection of an allomorph with a long vowel instead of one with a short vowel. I argue that these conditions are ordered and belong to their own designated module: lexical insertion conditions belong to Vocabulary Insertion and precede phonological ones which belong to Phonology. As such, theoretically, this article is an investigation into the mechanics of allomorphy selection within the framework of Distributed Morphology. The proposal is contrasted with existing proposals in the literature.

Empirically, the proposal allows one to understand variation in the distribution of allomorphs. More specifically, I discuss varying patterns in vowel length alternation in Dutch nominal and

verbal inflection and derivational affixation. I will argue that the distribution of the allomorph with the long vowel is often determined phonologically. The phonological patterns appear to be quite regular; when an appropriate allomorph with a long vowel is available and when its appearance would satisfy a phonological requirement to optimise syllabic structure, it will often be selected. Yet, the pattern is distorted by the lexicalized selection of allomorphs, which is, of course, by definition irregular. The article present a model which captures precisely this state of affairs and by doing so, it contributes to the field which struggled to reconcile precisely these two empirical patterns in a single model.

To be entirely clear, the term vowel length alternation refers to allomorphy and not to a phonological process. Given that stems with long and short vowels are alternating for a single lemma, one may have the impression that an underlying short vowel has lengthened and this is indeed what happened in Middle Dutch due to a process called Open Syllable Lengthening (see Lahiri & Dresher 1999 for a detailed discussion). Yet, vowel lengthening as a phonological process ceased to be productive in Dutch and what we merely see is that those stems that showed an alternation between a short and a long vowel due to the earlier phonological process acquired and kept this stem alternation. It is therefore improbable that the vowel actually lengthens in contemporary Dutch. It is more probable that two stem allomorphs, i.e. one with a short vowel and one with a long vowel, alternate.

The article is structured as follows. In the next section I discuss background literature. I give an overview on the state-of-the-art of theoretical proposals on the distribution of allomorphs. In Section 3 I illustrate the regular patterns of allomorphy in the domain of Dutch vowel length alternation. I show that these regular patterns follow from phonological principles that condition allomorphy to optimise syllable structure. Section 4 discusses irregular patterns of allomorphy which are ascribed to lexicalised uses of allomorphs. Section 5 discusses, for good measure, why certain patterns for allomorphy are excluded. The theoretical consequences of the data are discussed in Section 6. Section 7 sums up.

## 8.2 Existing theoretical proposals on the distribution of allomorphs

### 8.2.1 Introduction

In the empirical part of this paper, which is laid out in parts 3 and 4, it is argued that Dutch vowel alternating allomorphy is subject to both phonological principles and lexicalisation. As such, it presents a subset of the hypothetical factors that may determine allomorphy, which would also include syntactic contexts. The existing literature mainly focuses on syntactic and phonological factors and the current section presents this state-of-the-art of the theoretical and technical functioning of allomorph distribution.

As pointed out by Bonet and Harbour (2012: 220), there is little agreement in the literature as to which mechanism is responsible for the selection of an allomorph. Indeed, proposals range from stipulations or rankings at Vocabulary Insertion to phonological constraints in OT-based models. Bonet and Harbour point out that the source of the theoretical dispute probably stems from the disparity in the data (see also Bye 2007). Indeed, sometimes the distribution of allomorphs seems to be completely random, whereas other cases of allomorphy seem to follow phonological principles in a fully predictable and optimising fashion. Given this state of affairs, any approach that attempts

to capture all data by means of a unified account therefore seems to be doomed to be a less than elegant or satisfying account for at least some of the data.

On top of the fact that there is disagreement on which module should regulate the selection of allomorphs, there is the more basic question on how many modules are involved in allomorph distribution. The mere disparity of the data set might in fact indicate that the corresponding mechanisms and modules are disparate as well. The present paper follows this line of reasoning. In sections three and four I present a Dutch data set that indicates that a specific case of allomorphy shows simultaneously clearly optimising and predictable behavior and random idiosyncrasies. I argue that this state of affairs is actually indicative of the way allomorph selection is organised: it is split over two modules, resulting in two types of empirical patterns. More specifically, I argue that Vocabulary Insertion stipulates the lexicalised, idiosyncratic occurrences of allomorphs. In the absence of lexicalised selection, the competing allomorphs are sent to Phonology as disjunctive exponents of a single vocabulary item. Phonology will subsequently select the phonologically most optimal candidate. This position is supported empirically in later sections. The present section argues for this approach on theoretical grounds by reviewing the pros and cons of existing proposals.

### 8.2.2 Syntactic insertion contexts in Distributed Morphology

Distributed Morphology is a late insertion model; Syntax merges abstract features and all vocabulary insertion takes places post-syntactically at Vocabulary Insertion, which is an operation at PF. Hence, no exponents are present at Syntax. At the position of the root node (i.e. the lexical node), Syntax contains (at most) an index that refers to the relevant stem at Vocabulary Insertion, other nodes contain bundles of morphosyntactic features (Halle and Marantz 1993; 1994; Harley and Noyer 1999; Harley 2014).

When a syntactic cycle is complete, it is spelled out to PF and LF. At the PF-side, the structure may be subject to certain well-defined morphological operations before it is sent to Vocabulary Insertion. These morphological operations should not concern us here. Vocabulary Insertion is an operation that matches vocabulary items with the syntactic tree. The vocabulary items are acquired items that are stored in a list. At the LF-side, the compositional meaning of the syntactic features is computed. The compositionally derived denotation is combined with information from the Encyclopedia, another acquired list, which contains stored lexical semantics and instructions for (idiomatic) interpretation.

Since Distributed Morphology's seminal papers (Halle & Marantz 1993; 1994), the insertion of functional (i.e. non-root) vocabulary items is assumed to be subject to competition. It adheres to the Subset Principle, which is *in se* an Elsewhere Principle, giving priority to more specific rules over more general or default rules. These rules are often based on feature sets: vocabulary items are specified for the features they may realise. The vocabulary item of which the feature specifications are the best matching subset of the features on the syntactic node will be selected to realise the node. However, as allomorphs are mere variants of a single vocabulary item, they are usually specified for exactly the same features. As such, in the context of allomorphy, these rules typically do not depend on features, but on insertion contexts. More complex contexts are then given priority over less complex contexts. For example, the English tense suffixes  $\emptyset$ ,  $-t$  and  $-d$  cannot be

distinguished featurally as they all realise the same feature [+past]. What sets them apart is the type of verb with which they co-occur (example taken from Halle and Marantz 1993: 123-124):<sup>130</sup>

- (1) *Tns*
- [+past]  $\leftrightarrow \emptyset$  / [+strong] \_\_\_\_\_
  - [+past]  $\leftrightarrow /-t/$  / [+strong] \_\_\_\_\_
  - [+past]  $\leftrightarrow /-d/$

The exponent for which no context is given is the elsewhere form.

The insertion of lexical vocabulary items (also known as stems, roots or lexemes) was assumed to be subject to free choice rather than competition. As a result, competition between allomorphs was assumed to be restricted to the functional domain (Marantz 1997), as only functional vocabulary items compete for insertion in the first place. Instances of stem allomorphy were then postulated to result from readjustment rules, which can be understood as rules that guarantee some phonological readjustments of a stem at Spell-Out.

The theoretical distinction between functional and lexical allomorphy was motivated empirically by the observation that strong suppletion tends to be limited to the functional domain and stems only tend to show weak suppletion (see Marantz 1997 for discussion). However, Harley (2014) puts forward instances of strong suppletion amongst lexical verbs in Hiaki (see also Veselinova 2006). Having falsified the empirical distinction between allomorphy amongst functional and lexical vocabulary items, it becomes pointless to maintain the theoretical and technical distinction. Consequently, she proposes competition at Vocabulary Insertion for roots as well: lexical vocabulary items contain instructions to match an index with a phonological form, as illustrated in (2) (the example is based on Harley 2014). The example illustrates a Hiaki verb ('to run') that has two allomorphs, one of which is inserted in a plural context. The vocabulary item, which was merged at Syntax as an abstract index (e.g. the root with index 231) stipulates that the exponent *tenne* should be matched with this index in a plural context and with *vuite* in all other contexts:

- (2)  $\sqrt{231} \leftrightarrow / \text{tenne} / \quad / [\text{DP}_{\text{pl}} \text{_____}]$   
                   / *vuite* /       Elsewhere

As the example shows, vocabulary items may contain information on the availability and the distribution of allomorphs. What may determine the use of an allomorph is a stipulation on the vocabulary item that a certain allomorph is the more specific candidate for a certain context. For the example above, the selection mechanism is the stipulation that *tenne* is a closer match for plural contexts. Note that the two allomorphs spell out the same index and their selection solely depends on an insertion context. Note further that the so-called *syntactic* selection of allomorphs is strictly speaking not determined by the module Syntax itself: allomorphy is determined at Vocabulary Insertion. It is *syntactic* in the sense that Vocabulary Insertion takes syntactic contexts into account.

To summarise, in Harley (2014) lexical selection depends on features for functional items and on an index for roots. Once the relevant vocabulary item is selected, the selection amongst its

---

<sup>130</sup> I merely adopt the example from Halle and Marantz (1993) to illustrate their views on the technical distribution of allomorphs. Not everyone would endorse the view that *-t* and *-d* are allomorphs for example: Kayne (2016) derives them both phonologically from an underlying *-t*. However, this discussion should not concern us here.

various allomorphs (if any) depends on insertion contexts stipulated at the vocabulary item. Richer insertion contexts precede less specified ones. In other words, allomorph selection still functions as in Halle and Marantz (1993), albeit extended to root allomorphy.

### 8.2.3 Allomorphs spelling out a syntactic head

In the previous section we have seen examples of allomorphs that depend on a given syntactic context, without assuming that the allomorph indeed spells out a specific syntactic feature itself. Yet, this possibility also seems to exist. It is, for example, well-known in the literature that syntactic plurality may trigger allomorphs or other irregular exponents. Yiddish, for example, shows allomorphy in plural contexts (Lowenstamm 2007). In the following example the root *kind* 'child' alternates with its allomorph *kinder*:

- |     |         |            |                |                   |
|-----|---------|------------|----------------|-------------------|
| (3) | a. kind | b. kinder  | c. kind-l      | d. kinder-l-ex    |
|     | child   | children   | child-DIM      | child-DIM-PL      |
|     | 'child' | 'children' | 'little child' | 'little children' |

An irregular exponent may thus be found in syntactically plural contexts. There is a strong tendency in the literature to analyze these irregular markers as instances of noun class markers of plurality. It is often proposed that the noun class marker is an instance of  $n^\circ$ , which merges above the root and which is selected by the plural context. Versions of such a proposal have been endorsed by Lowenstamm (2007) for Yiddish, by Kramer (2009) for Amharic and by Acquaviva (2008) for Breton. In the same vein, Hermans & Postma (2009) have proposed that Dutch vowel length alternation and a Dutch allomorph pattern resembling the Yiddish data in (4) are instances of noun class marking in plural contexts as well, a proposal to which I briefly come back in Section 5.2. In all these proposals, the alleged irregularity of the stem is in fact the spell-out of a syntactic head, viz. a noun class marker. It is thus argued in the literature that a difference in form in such cases reflects difference in structure. Hence, it is referred to as *pseudo-allomorphy*.

Technically, pseudo-allomorphy can be captured in three ways. Firstly, it can be argued that the instances should be aligned with the cases presented in Harley (2014): the allomorphs are systematically triggered in the presence of a given syntactic head and, as such, they *seem* to spell out this head, but technically, they do not. They simply systematically appear next to this head, which may be realised by a silent exponent in itself. Secondly, it can be argued that the allomorph is technically not an allomorph, but a concatenation of two lexical items. For example, under such a proposal *kinder* is not an allomorph of *kind*, but the spell-out of *kind* and *-er*. Under such a proposal, these data do not instantiate allomorphy and they become irrelevant to the discussion. Thirdly, one might argue that a form such as *kinder* indeed spells out both the root and little  $n$  simultaneously. Given that *kind* and *kinder* then spell out different feature sets, their selection is subject to the subset principle. Such proposals have to adopt a version of phrasal spell-out and are incompatible with approaches in Distributed Morphology that adhere to the original idea of piece-based realisation. I will come back to this idea in Section 2.8.

A thorough discussion of this issue lies beyond the scope of the present paper. It will become clear that I argue that the data presented in this paper are not to be analysed as syntactic allomorphy.

#### 8.2.4 Lexicalised insertion contexts

Lexicalisation may be responsible for allomorphy selection. For example, the participle of the verb *uitverkiezen* ‘to select’ is *uitverkozen*, as can be seen in (4):

- (4) Jij bent uitverkozen.  
you are selected  
'You are selected.'

Yet, there is an idiomatic reading in which a different form is used:

- (5) het uitverkoren volk  
the selected people  
'the Chosen People (i.e. the Jewish people)'

There is nothing systematic to be mentioned about the exceptional use of *uitverkoren* in the example above. A connection has been stored between an irregularity in form and an irregularity in meaning in a specific collocation. The use of an irregular form has simply been listed as the conventionalised form to express a specific idiomatic meaning. The selection of the allomorph is then an arbitrary convention, a conventionalised irregularity. An allomorph may thus be selected in a specific context despite the fact that there are no principled, i.e. phonological or syntactic, grounds to do so (see also Section 2.5).

The mechanics presented in Harley (2014) on syntactic allomorphy can be straightforwardly extended to incorporate such examples of lexical allomorphy. Just as syntactic insertion contexts can be stipulated on a vocabulary item, so can lexicalised insertion contexts. In fact, there is a good reason to group them technically. The information both types of insertion contexts rely on can be read off from the syntactic tree. This is obvious for allomorphy based on syntactic insertion contexts. But also for lexical contexts the tree does contain all the necessary material to recognise a relevant collocation, as its nodes contain the indices or features of the other lexical items referred to. Furthermore, both for lexical and syntactic insertion contexts, no systematic rule is at work (beyond locality). The distribution of the allomorph simply depends on a listing associating the use of a specific allomorph with neighbouring material in the tree.

#### 8.2.5 Rich subcategorisation frames

As described above, Harley (2014) argues that the syntactic selection of an allomorph can be understood technically as a stipulation on the vocabulary item, determining in which syntactic context a certain allomorph should be inserted. *Mutatis mutandis*, if the selection of an allomorph is phonological, one could stipulate a phonological context on the vocabulary item, as in the following abstract example:

- (6)  $\sqrt{\text{index}} \leftrightarrow \begin{cases} \text{allomorph 1} & / \text{specific phonological context} \\ \text{allomorph 2} & \text{Elsewhere} \end{cases}$

Again, in such a scenario it is not the module Phonology in itself that may select the allomorph, it is rather a phonological context stipulated on the vocabulary item that will ensure the insertion of the phonologically more optimal candidate.

An approach such as the one sketched here is presented in Paster (2006). Discussing suppletive forms of affixes, she proposes that all conditions determining allomorphy are simply part of the subcategorisation frame of the vocabulary item, *in casu* the affix. As a result, vocabulary items in her model may have an elaborate subcategorisation frame, including both syntactico-semantic and phonological insertion criteria (see, for example, Paster 2006: 12).

From a descriptive point of view, such subcategorisation frames successfully capture all the facts. It suffices to adequately list all insertion contexts with an appropriate ranking. Model-wise, the proposal dovetails in general with Distributed Morphology, as it extends ideas already present in the seminal papers (cf. example 1 in Section 2.2). However, even though, technically, the stipulations on the vocabulary items will derive the desired results, we may be reluctant to adopt these adjustments as the final say on the organisation of vocabulary insertion for two reasons. Firstly, stipulating the relevant phonological context on the vocabulary item does capture the observation that a specific form will be inserted in a specific phonological context, but it does not express the insight that such phenomena are often motivated. In fact, one could stipulate just anything (that respects locality conditions) as an appropriate phonological context for insertion for a given allomorph. In this sense, it is a brute force solution. The insight that the insertion of an allomorph often actually happens to optimise the phonological make-up of the word is lost and not expressed by the current mechanism.

Admittedly, according to Paster (2006) this result is actually an advantage of the approach. She points out that not all phonologically-driven cases of allomorphy are actually optimising. She presents twelve examples from various languages in which the distribution of suppletive forms of affixes is clearly determined by phonological properties of the lexical item involved (Paster 2006: 76-98). Yet, the use of these suppletive forms is at best neutral; their occurrence does not optimise the word phonologically.

Given the existence of these facts, note the following. As rightfully pointed out by Paster, any account dealing with the distribution of allomorphs has to be able to capture phenomena that are irregular, arbitrary or lexicalised in nature. A theory in which the most optimal candidate invariably wins is simply empirically inadequate. Note that it follows in particular, as Paster (2006) points out, that trying to capture phonologically motivated types of allomorphy by solely relying on Optimality Theory will therefore be an ordeal, given that it is easy to find non-optimal instances of phonological allomorphy, a point to which I come back below.

Nevertheless, Paster's approach asks us to throw the baby out with the bathwater. There is no contradiction in fully recognising the existence of counterexamples, which certainly should be captured one way or another, and adhering to the insight that a general phonological principle may determine other instances of allomorphic distribution. There are cases in which phonologically conditioned allomorphy does show regular optimising behavior and Paster's (2006) subcategorisation-frame based approach does away with the tools to capture them, given that all phonologically motivated suppletion is technically reduced to idiosyncratic properties of vocabulary items.

The second reason because of which we may be reluctant to adopt Paster's (2006) approach fully is the fact that phonologically-driven allomorphy may show uniform behaviour over various vocabulary items. From a phonological point of view, this phenomenon is not hard to understand; all vocabulary items that share a relevant phonological feature may be subject to the same phonological condition. Yet, under a subcategorisation-frame based proposal, such a natural grouping becomes a coincidence. The fact that the linguist understands why certain vocabulary

items may show similar behaviour cannot be expressed in a model that merely relies on stipulations on individual lexical items.

In sum, subcategorisation-frame based proposals have strong descriptive force; all conditions determining allomorphy can be stipulated on the lexical item. Furthermore, it does recognise and capture the existence of non-optimising counterexamples. The disadvantage is that these stipulations are indeed mere stipulations, which fail to express insights. Another disadvantage is that the approach is lexical item based, which fails to express patterns generalising over various lexical items. It appears that the approach has not much to offer when it comes to phonologically driven patterns, but it does offer us the tools to deal with exceptions.

### 8.2.6 Phonological rules as insertion contexts

Just as Paster (2006), Nevins (2010) stipulates phonological conditions for allomorphy selection at Vocabulary Insertion. Yet, he takes a slightly different approach. He simply specifies the phonological rule as a contextual condition for insertion on the vocabulary item, as in the following example (taken from Nevins 2010: 22, his example (38)):

- (7) INDEF ↔/æn/ if it removes a violation of ONSET

The advantage of incorporating a rule rather than a phonological context is clear; the vocabulary item expresses an insight rather than an arbitrary mechanism. It may seem like we have addressed the concern.

However, consider the following problem. If one allows both syntactic and phonological conditions to be specified on the vocabulary item, we define Vocabulary Insertion as a hybrid operation. It should be able to read both syntactic and phonological conditions simultaneously. In this respect the stipulation that an allomorph may be sensitive to a specific phonological context, as in Paster (2006), is the more conservative stipulation; Vocabulary Insertion simply must be able to check surrounding syntactic and phonological features. However, if we adopt Nevins' proposal, Vocabulary Insertion must be able to actually interpret a phonological rule. It defines Vocabulary Insertion as an operation that is able to execute quite some phonology. Yet, simultaneously, we want it to be able to check for adequate syntactic conditions as well, following Harley's proposal. In sum, it defines Vocabulary Insertion as a two-headed dragon.

Note, furthermore, that a single lexical item may be subject to both insertion contexts and phonological stipulations. Such stipulations are ranked according to the Elsewhere Principle and one of them will precede the other one. Now, given that stipulations are still related to lexical items, the approach is ill-fitted to capture more general patterns over various lexical items. More in particular, we will see that insertion contexts systematically win over all phonological constraints (see Sections 2.8 and 2.9). This systematic ordering over various lexical items are technically nothing but instances of coincidence.

Finally, one might consider the fact that phonological constraints are ideally understood as universal constraints. Vocabulary Insertion, in contrast, is a list of acquired information. As such, it is odd to assume that Vocabulary Insertion hosts phonological principles.

In sum, it is advantageous to incorporate a phonological rule rather than a phonological context as an insertion context. However, it implies that Vocabulary Insertion can execute phonology and that it hosts both innate and acquired insertion principles. The relation between Vocabulary Insertion and Phonology proper becomes unclear: if Vocabulary Insertion can handle Phonology, is

it then defined as the module Phonology or does it still co-exist with a separate module called Phonology?

### 8.2.7 Proposals within Optimality Theory

Optimality Theory (McCarthy and Prince 1993) defines morphological and phonological conditions for allomorphy selection as constraints. As such, morphological conditions and phonological conditions are evaluated in the same module. OT thus relies on a hybrid module as well. Older work in OT assumes that these conditions are ordered: one can formulate a general constraint that stipulates that all phonological conditions are to precede all morphological requirement (McCarthy and Prince 1993) or the other way around (Yu 2003: 108).

There is of course the well-known general problem that it is hard to see how these original proposals can be adopted by a realisational model such as DM (and see Paster 2006 for a thorough critical discussion). Distributed Morphology cyclically matches vocabulary items to syntactic nodes from the root outwards and the match is evaluated for each node within each cyclic step by the Elsewhere Principle. As such, it is principally incompatible with a model that maps a surface representation to an underlying representation for a phonological word without taking into account its internal structure (as discussed at length in Embick 2010).

Apart from this theoretical objection, there are two empirical objections. Firstly, as pointed out by Paster (2006: 77) in the domain of phonological allomorphy non-optimising examples can be found easily. They are challenging to OT, to say the least. Secondly, OT evaluates input candidates to derive surface forms. There are no grounds for underlying representations to trigger phonological allomorphy. Yet, underlying representations may trigger allomorphy, as will become clear in Section 3.3.

Wolf (2008) presents a serial version of OT to deal with phonological allomorphy (Candidate Chains, McCarthy 2007). In such an approach, candidates may improve incrementally. As such, the theory opens up the possibility to integrate underlying requirements into intermediate steps. Wolf's main proposal is that morphological insertion criteria and phonological insertion criteria are evaluated in the same module, i.e. phonology. Wolf rewrites the subset principle as a series of constraints that guarantee that the candidate of which the features are the closest subset of the syntactic node will outrank less optimal matches. However, unlike the older versions of OT, he does not assume that there is an intrinsic ordering between constraints guaranteeing feature matching and those guaranteeing the application of phonological principles: morphological and phonological constraints can interleave. Just as Nevins (2010), he thus assumes that a single module hosts both universal phonological principles and acquired information, i.e. the link between exponents and the feature sets they realise.

Now, it is important to note that the discussion in his thesis is limited to two particular types of allomorphy. Firstly, he addresses instances of true phonological allomorphy, i.e. data which can be fully captured on phonological grounds as they show regular, optimising behavior. Secondly, he presents instances of phonologically motivated lexical item selected which, arguably, would not even be called allomorphy in Distributed Morphology. For example, he discusses the distribution between the determiners *el* and *la* in Spanish. As is well-known, *la* is selected in feminine NPs, *el* is selected in masculine ones. However, *el* may also be selected by feminine nouns that begin with stressed [á]. Clearly, the distribution between *el* and *la* is thus partially phonologically motivated. However, strictly speaking, *el* and *la* may not be allomorphs of a single lexical item, but simply two lexical items competing for insertion. This example characterises his approach: allomorphs are

distinguished featurally and the selection is then done by a morphological constraint that will select the closest matching feature set.

Wolf is forced to invariably assume featural distinctions between allomorphs. A first problem is pointed out by Bonet & Harbour (2012: 227): the featural distinction may be entirely stipulative. The second problem is that the approach is at odds with the essence of how allomorphy is understood in Distributed Morphology. Recall from Section 2.2 that allomorphy in DM is not understood as competition based on feature sets; it is based on ranked insertion contexts. Wolf does integrate Vocabulary Insertion in Phonology by rewriting the subset principle as constraints, but he remains silent on those data that are a central aspect to the understanding of allomorphic distribution in DM, i.e. the ranking of insertion contexts. In sum, it is unclear how Wolf's proposal can be applied to instances of allomorphy that are not to be distinguished featurally, as discussed in Section 2.2 and 2.4.

The empirical domain discussed in this article does not match Wolf's approach very well. We will see that allomorph selection for Dutch vowel alternating stems is both based on phonological principles and on stored, lexicalised information. Because some of the data are lexicalised, rather than being based on a selection mechanism relying on syntactic features, the data are at odds with Wolf's proposals. I will therefore not come back to this approach in later sections.

### 8.2.8 Split proposals

Booij (1998) discusses instances in Dutch morphology where phonological principles clearly seem to regulate the distribution of strong and weak allomorphs, but the application of these phonological principles are not without exception. For example, he discusses the by now well-known fact that the two competing Dutch plural affixes *-s* and *-en* are subject to syllabic well-formedness: *-en* is favoured if it transforms the root into a trochee (with right directionality). Hence, the Dutch plural of *kat* 'cat' is *katten*, but the plural of *dokter* 'doctor' is *dokters*. Even though this tendency is undeniably present in Dutch, counterexamples and examples of regional variation are not hard to find. For example, as Booij points out, speakers in the Netherlands favour the phonologically less optimal plural form *tests* 'tests' for the noun *test*, whereas Belgian speakers will use the phonologically expected form *testen*. Speakers in the Netherlands simply conventionalised the unexpected form.

To account for such examples, Booij (1998) proposes that stems and affixes are marked with subcategorisation features which may be of any kind (phonological, syntactic, ...). These features regulate morpheme concatenation in morphology. Morpheme concatenation acts as a generator that generates possible candidates which are then sent to Phonology for evaluation in an OT fashion. As subcategorisation features precede phonological principles in the proposal, Booij can guarantee that phonology will operate within the language's morphological boundaries and he can make sure that phonology -and nothing but phonology- indeed takes place at a designated module, viz. Phonology.

*Mutatis mutandis*, the proposal can be adopted in Distributed Morphology. Subcategorisation features can be understood as being part of Vocabulary Insertion, as in Paster (2006) and Harley (2014). The output of Vocabulary Insertion, which can include various candidates for the spell-out of a single node, are then shipped to Phonology for further evaluation. It will become clear in the next section that the present proposal adopts exactly this idea.

Bye and Svenonius (2010; 2012) formulated a proposal that shows similar features, but which differs in important aspect. They propose to split Vocabulary Insertion into two different operations, which they call Match and Insert. Match precedes Insert; Match's output is Insert's input.

Match can see syntactic features and syntactic structure, such as dominance, movement and phase boundaries and it matches lexical items to the syntactic tree. As in Distributed Morphology, it is assumed that those lexical items are chosen that realise a maximum of features present in the tree. Domains for insertion are not terminal nodes, as in Distributed Morphology, but rather Spell-out domains, i.e. phases. As such, a lexical item may realise a span of terminal nodes (cf. Williams 2003).<sup>131</sup> Such a span is understood as a segment of an extended projection. For example, the noun plus number marking is understood as a phase and thus may be realised as a span in the noun's extended projection, i.e. the DP. In the English noun phrase, the noun will incorporate into the plural head, due to the fact that English number marking is suffixal. This type of linearisation is characteristic of the proposal: all linearisation that can be done in syntax will be done in syntax. The suffixal nature of the English plural thus does not need to be stipulated as a property of the suffix, as it is already derived at syntax. This cyclic domain is sent to Match, where a portmanteau morpheme, such as *geese* and *mice*, or several lexical items, such as *ballerina-s*, may realise this span. A form such as *mice* will be given preference over a form such as \**mouses* or *mice-*Ø due to a principle called *Minimize Exponence*: using less exponents to match the syntactic features are better than using more exponents. Note that the model thus has a mechanism to capture allomorphs that disjunctively realise different syntactic feature sets: *mouse* and *mice* are allomorphs and their distribution is regulated by postulating that they realise different features. The model allows for a different type of allomorph selection as well: the model allows various disjunctive forms to be selected by Match. The choice between these exponents is postponed. As such, allomorph selection can be assigned to a later module.

Match's output are candidates for insertion and are sent to the module Insert. Insert has the following tasks: it selects allomorphs on phonological grounds in case of disjunctive exponence, it linearizes the structure and it builds an underlying phonological representation. Of particular interest to us, is that the module thus contains an allomorph selection mechanism. Bye and Svenonius (2010: 14) provide the following example. The lexical item of the English indefinite article contains the following disjunction:

- (8) {ən} / \_\_ V ↔<INDEF>  
       {ə}              ↔<INDEF>

Given that the choice between *an* and *a* is based on phonological criteria, the selection is not established by Match, which can only take into account syntactic criteria. The disjunctive candidates are thus both sent to Insert, which picks the best candidate on phonological grounds.

An important advantage of the split approach presented here is that the insertion mechanism is able to take both syntactic and phonological requirements into account, while it does not require a single operation to interface with both syntactic and phonological insertion contexts. However, note that the phonological selection in (9) is based on an insertion context rather than on a principle. This is a missed opportunity. In Section 2.6 we argued that if a phonological selection can

---

<sup>131</sup> In this respect, the approach differs from Distributed Morphology, in which the spell-out of terminal nodes is assumed. Approaches that rely on spans can be reformulated into a terminal node approach by relying on fusion and null affixes and thus can be made compatible with Distributed Morphology.

be formulated as a principle, one misses a generalisation by reducing it to a context. We thus end up with a separate module for phonology that does not really execute phonology with full force.

Bye (2007) explicitly discusses allomorphy in yet another split approach. He concurs with Paster (2005) that many examples of allomorphy are not phonologically optimal. He points out that these examples, which he labels *morpholexical* are language-specific and inviolable. If these rules are not respected, they lead to ungrammaticality. In contrast, phonologically selected examples of allomorphy are selected on the basis of universal, violable principles that guarantee the selection of an optimal candidate. According to Bye the process of Eval precedes a Morpholexical Control mechanism: the output of phonology is checked by Control and results that are not compatible with language-specific morpholexical criteria are rejected. Bye's (2007) split approach model thus succeeds in capturing both phonological and morpholexical selection. The mechanics of his proposal fit the data: phonological selection is done in a *bona fide* phonological module that operates on the basis of universal phonological principles. Morpholexical selection, on the other hand, works on the basis of language-specific, listed and inflexible information. Bye (2007) thus presents a model that successfully and strongly captures the disparate data. He argues to acknowledge that both types of data exist and he gives them an appropriate place in the model. The present article aims to contribute to the same goal. Bye himself points out that a possible objection against the model might be that OT is complemented with a set of declarative constraints (in the control mechanism). In the following section I propose not to adopt the control mechanism. I argue that it suffices to let Vocabulary Insertion precede Phonology (as we assume in Distributed Morphology anyway) to accommodate for both data and both selection mechanisms.

### 8.2.9 Candidate sets

Let us take stock of the data and the ideas presented in the previous sections. Empirically, we have encountered four different types of allomorphy: some allomorphs are sensitive to a syntactic context (Section 2.2), some are sensitive to a lexical context, which is possibly phonologically non-optimal (Section 2.4), some allomorphs disjunctively realise different syntactic feature sets (Section 2.3 and the discussion on Wolf 2008 in Section 2.7) and some differ qua phonological optimality (Section 2.7). Approaches that attempt to unify allomorphy into a single module will either ignore some of the data and be observationally inadequate (cf. the critique in Paster 2006 and Bye 2007 on Wolf 2008) or they will have to rely on rich subcategorisation frames (as in Paster 2006), which is a brute force solution to achieve observational and descriptive adequacy.

Now note that three of the four types of allomorphy can be captured by Vocabulary Insertion. We have seen that Vocabulary Insertion sees information present in the tree and it can deal with disjunctive feature sets and disjunctive insertion contexts via the Elsewhere Principle. As such, it can successfully deal with allomorphs realising different syntactic feature sets and it can take into account allomorphy sensitive to a syntactic context. Given that Vocabulary Insertion is in itself a list of lexical items, it also can see lexicalised information stipulating stored irregularities on those lexical items. Such lexical items will have the following, familiar format:

Let us thus assume that Vocabulary Insertion occupies its familiar spot in the Y-model and that it regulates the selection of these three types of allomorphy, relying on no other basic mechanisms than those that were already proposed in Distributed Morphology's seminal papers.

I now adopt the idea proposed by Bye and Svenonius (2010) that the output of a Vocabulary Insertion mechanism may be a set of candidates. To do so, we do not have to alter our assumptions on Vocabulary Insertion drastically. We simply have to allow a tie between candidates that seem to be equal options at Vocabulary Insertion and which can only be distinguished on the basis of phonological criteria. It is a minor modification of Vocabulary Insertion as we know it. I propose that a lexical item may be marked for a set of unranked exponents, i.e. allomorphs, at Vocabulary:

- (10)  $\sqrt{539} \leftrightarrow \{\text{allomorph}_1, \text{allomorph}_2\}$

Such a situation will only occur if there are no lexical or syntactic criteria to distinguish between the allomorphs. This candidate set is then shipped to Phonology, which will select the best candidate on the basis of universal phonological constraints. The possibility of lexical items, such as the one proposed in (11), which may exist alongside familiar lexical items, such as the one in (10), is basically the only innovation here, but its effect is considerable. Note that the fact that the allomorphs are not ranked is not a problem. Stipulating an elsewhere form is not necessary, given that one of the forms will always be more optimal than the other one phonologically.

It is important to note that nothing prevents us from mixing the ingredients presented in (10) and (11). As such, a morpholexical stipulation at the vocabulary item can block a richer set of candidates right from the start, as illustrated in the abstract example in (11):

- (11)  $\sqrt{539} \leftrightarrow \begin{array}{l} \text{allomorph}_1 / \text{specific insertion context } \alpha \\ \{\text{allomorph}_1, \text{allomorph}_2\} \end{array}$

(12) says that one allomorph will always be inserted in context  $\alpha$ , in all other contexts the decision is up to Phonology. We have seen that any approach to allomorphy selection should be as successful in capturing the phonological patterns as in capturing the pervasive irregularities. As such, having a mechanism that can give priority to exceptions is an important addendum.

In sum, the purpose of this article is to defend a split approach to allomorph distribution which sets apart allomorphy based on stored insertion contexts from allomorphy based on phonological principles, and as such it argues against proposals that group all criteria together in Vocabulary Insertion (e.g. Paster 2006) or Phonology (e.g. Wolf 2008). It argues that both types should be assigned to their own designated module: insertion contexts are part of Vocabulary Insertion, phonological criteria belong to Phonology. As such, the approach takes advantage of the strength of both modules: Vocabulary Insertion, by definition a stored list, captures irregularities and stored insertion contexts, Phonology, based on universal constraints, provides for phonologically optimising patterns. Furthermore, these modules are ordered: Vocabulary Insertion precedes Phonology. The main reason to assume this order is the logic of the Y-model. Vocabulary Insertion interfaces with the morphosyntactic features and indices present in the syntactic tree. Hence, its input must be the syntactic tree. Phonology, on the other hand, derives a surface representation from an underlying representation, hence it must be fed the underlying representation of the exponents, which follows immediately if it follows Vocabulary Insertion.

A proposal on the exact nature of Phonology lies outside the scope of this article beyond the assumption that it is some kind of optimising model (see Bye 2007 for a discussion). It will become clear that the data in the present article suggest that also underlying representations can be subject to phonological principles determining allomorph distribution. This criterion is a desideratum for a phonological module in my opinion.

In the following sections I discuss a data set from Dutch that shows how the proposed model is beneficial. The field has gone back and forth on concentrating on either irregular, phonological non-optimal allomorphy or fully phonologically optimising allomorphy, rather than acknowledging that both instances co-exist (as emphasised in Bye 2007). Dutch contains a data set which illustrates exactly this point: within a single data set, both patterns clearly co-exist.

### 8.3 The fully regular phonological patterns

The objective of the empirical part of the paper is to show that there are vowel alternating stems in Dutch that are simultaneously subject to a fully regular, optimising pattern and to stored stipulations. The current section presents the first half of the facts: it discusses the phonological motivation to select an allomorph with a long vowel. It presents those syntactic domains in which the phonological rule is applied without any exception, viz. nominal number inflection and verbs.

#### 8.3.1 The phonological motivation for vowel length alternation

Van der Hulst (1985) contains a proposal on why vowel length alternation might optimise Dutch syllable structures. Consider the following examples of vowel length alternation in Dutch from the domain of derivation (Van der Hulst 1985: 63). They show an alternation between stems with a short vowel in morphologically simplex contexts and forms with a long vowel in derived forms.

- |      |    |        |     |          |            |     |             |
|------|----|--------|-----|----------|------------|-----|-------------|
| (12) | a. | demon  | [ɔ] | 'demon'  | demonisch  | [o] | 'demonic'   |
|      | b. | satan  | [a] | 'satan'  | satanisch  | [a] | 'satanic'   |
|      | c. | Israel | [ɛ] | 'Israel' | Israelisch | [e] | 'Israeliic' |

Van der Hulst (1985) provides a phonological account: he proposes that the stem with the long vowel is selected to avoid ambisyllabicity. The proposal goes as follows. A Dutch minimal rhyme is at least bimoraic. The hypothetical forms in (13) are therefore excluded by Dutch phonology:

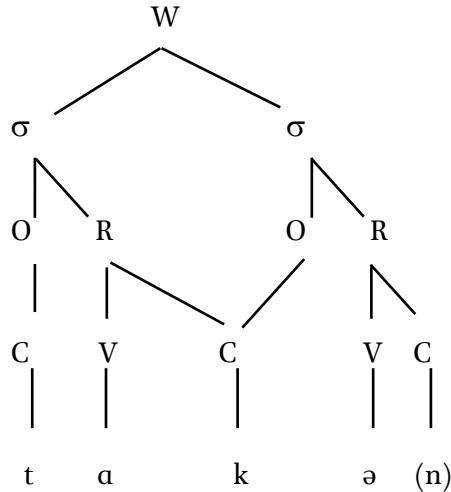
- |      |    |       |
|------|----|-------|
| (13) | a. | *[nɔ] |
|      | b. | *[da] |

As a result, short vowels create a context of ambisyllabicity in Dutch. To avoid a rhyme which is monomoraic, Van der Hulst (*ibidem*) proposes that a consonant in the onset of a following syllable may function simultaneously as the coda of the previous syllable. In other words, the consonant may be ambisyllabic. This is illustrated in the example in (15) in which the /k/ is shared due to the fact that the short vowel /a/ only occupies a single moraic position.

- (14) tak-en ['takə(n)]

branch-PL

'branches'



Ambisyllabicity, of course, does not occur in closed syllables or if the rhyme of the first syllable contains a long vowel. Both closed syllables and long vowels occupy two moraic positions. As such, they fulfil the requirements of the minimal Dutch rhyme on their own.

Van der Hulst (1985) does not only argue for ambisyllabicity, he also proposes that spreading of the vowel is preferred in Dutch over spreading of the consonant. In other words, vowel length alternation is preferred over ambisyllabicity. Hence, if a stem allomorph with a long vowel is available for the root, this allomorph will be selected to fulfil the phonological requirement of the minimal rhyme. Whether such an allomorph is available is a stored property of the root. This leads us to the following phonological rule:

- (15) i) *Van der Hulst's ambisyllabicity:*

To avoid a rhyme which is monomoraic, a consonant in the onset of a following syllable may function simultaneously as the coda of the previous syllable.

- ii) *Van der Hulst's vowel spreading, adapted.<sup>132</sup>*

To avoid a rhyme which is monomoraic, a stem allomorph with a long vowel may be selected, if available.

- iii) *Van der Hulst's ordering:*

Vowel spreading is preferred over ambisyllabicity.

Van der Hulst's proposal shows similarities with Lahiri & Dresher's (1999: 706-710) account of the productive process of Open Syllable Lengthening in Middle Dutch.<sup>133</sup> They propose that there was a pressure to maximize the main stressed foot. In other words, monomoraic stressed open syllables, i.e. open syllables with a short vowel, have shifted to bimoraic open syllables, i.e. open syllables with a long vowel. The authors point out that stressed light syllables have to be bound together with a second syllable in order to acquire the needed second mora as the head of a foot must dominate at least two morae. It is not pointed out explicitly by the authors, but an immediate side-effect of the

<sup>132</sup> Strictly speaking, Van der Hulst (1985) argues in favor of vowel length alternation, but he does not mention the fact that it may go back to stem allomorphy.

<sup>133</sup> Yet another account can be found in Lahiri & Dresher (1991). As Lahiri and Dresher themselves reject the proposal in favor of Lahiri & Dresher (1999) (see Lahiri and Dresher 1999:fn47), I limit the present discussion to the more recent work.

pressure to maximize the main stressed foot is that they no longer require being bound to the second syllable. As such, the account is similar to Van der Hulst's proposal.

Lahiri & Dresher (1999) and Van der Hulst (1985) thus have a similar process in mind, the main difference being that Lahiri & Dresher (1999) formulate the principle as motivated by the foot, preferring a bimoraic head, whereas Van der Hulst (1985) formulates it as a requirement of the syllable, preferring vowel lengthening over ambisyllabicity. Lahiri & Dresher's (1999) proposal was formulated for Middle Dutch, but it is interesting to check whether it fares better for contemporary Dutch than Van der Hulst's (1985) proposal. Both proposals make a different empirical prediction when it comes to vowels in open syllables, followed by a second syllable, in non-stressed positions in those stems that show the relevant vowel alternating allomorphy. In short, we should look at words in which the relevant vowel occurs in an environment that is subject to the syllabic configuration relevant to Van der Hulst's proposal, without being the head of the foot, which is the main criterion in Lahiri & Dresher's (1999) proposal. Such contexts can be created by adding stress-bearing suffixes to those nouns in which the relevant vowel is not in a stressed position, as in (17). The suffixes serve to provide a second syllable in the configuration. In the examples, main stress is indicated by a superscripted accent, secondary stress is indicated by a subscripted accent, the vowel that changes under allomorphy is given in bold. The secondary stress corresponds to the main stress of the stem when it would not bear a suffix. The transcriptions are based on Heemskerk & Zonneveld (2000).

- (16) a. professoraal [prɔ̄fessor'ɑl] 'professoral'
- b. professoraat [prɔ̄fessor'ɑt] 'professorate'
- c. satanisme [satən'ɪsmə] 'satanism'
- d. pastoraal [paſtɔ̄r'ɑl] 'pastoral'
- e. pastorie [paſtɔ̄r'i] 'vicarage'
- f. demonie [demōni'ɪ] 'demonism'
- g. demonisme [demōni'ɪsmə] 'demonism'

Despite the absence of stress, the stem allomorph with the long vowel does appear in the examples in (17), even though the vowel is not the head of the foot. It thus appears that the relevant syllabic context suffices to trigger the allomorph with the long vowel. Given the data in (16), it seems that Van der Hulst's proposal has better empirical coverage for contemporary Dutch than Lahiri & Dresher's (1999) proposal (which has been formulated for Middle Dutch). As such, we will proceed with Van der Hulst's proposal as formulated in (15).

It is clear that Lahiri & Dresher's (1999) account of the Middle Dutch facts and Van der Hulst's (1985) account of Dutch are similar in nature. It is interesting to note that, as such, it is reasonable to assume that the underlying prosodic pressure that motivates the stem alternation in contemporary Dutch is probably the same as the one that motivated the original productive phonological process in Middle Dutch. Dutch may have lost a phonological process, but it did not lose its settings on what constitutes an optimal prosodic configuration and we still see it kicking in as soon as it has a stem allomorph to satisfy its demands.

### 8.3.2 The phonological pattern for nominal number inflection

Standard Dutch has two plural markers, viz. *-en* and *-s*. In general, they do not give rise to any changes in the stem. This is illustrated in (17)-(20):

- |                       |   |
|-----------------------|---|
| (17) a. een kat [kat] | b. twee kat-en [katə(n)] <sup>134</sup> |
| a cat                 | two cat-PL                              |
| 'a cat'               | 'two cats'                              |
- 
- |                        |                           |
|------------------------|---------------------------|
| (18) a. een zoon [zon] | b. twee zoon-en [zonə(n)] |
| a son                  | two son-PL                |
| 'a son'                | 'two sons'                |
- 
- |                          |                         |
|--------------------------|-------------------------|
| (19) a. een otter [ɔtər] | b. twee otter-s [ɔtərs] |
| an otter                 | two otter-PL            |
| 'an otter'               | 'two otters'            |
- 
- |                      |                     |
|----------------------|---------------------|
| (20) a. een oom [om] | b. twee oom-s [oms] |
| an uncle             | two uncle-PL        |
| 'an uncle'           | 'two uncles'        |

However, there are around thirty to forty irregular nouns which do show a stem alternation (see also Booij 1995: 72). The root's vowel lengthens in the plural and it may lengthen in the diminutive. An example is given in (21):

- |                       |                           |
|-----------------------|---------------------------|
| (21) a. een vat [vat] | b. twee vaat-en [vatə(n)] |
| a barrel              | two barrel-PL             |
| 'a barrel'            | 'two barrels'             |
- 
- |                        |                            |
|------------------------|----------------------------|
| c. een vaat-je [vat̪ə] | d. twee vaat-je-s [vat̪əs] |
| a barrel-DIM           | two barrel-DIM-PL          |
| 'a small barrel'       | 'two small barrels'        |

I mention the diminutive forms here for good measure, but I postpone a discussion of these forms till Section 4.1.

For the relevant stem alternating nouns, the vowel length alternating pattern for number inflection in the non-diminutive forms is fully regular: in non-diminutive forms singular nouns select the allomorph with the short vowel, plural nouns select the allomorph with the long vowel. The data thus allow for the following empirical generalisation:

- (22) If, for a Dutch stem, both an allomorph with a short vowel and one with a long vowel are available, the allomorph with the short vowel will be selected in a non-diminutive singular NP

---

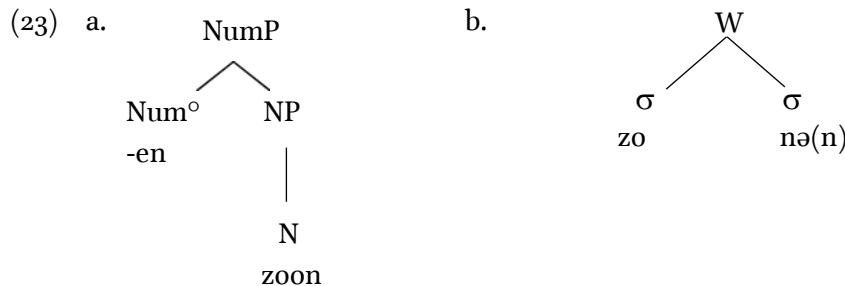
<sup>134</sup> The pronunciation of the final /n/ of the plural marker is optional and depends mostly on the speaker's dialect and the phonological context.

and the allomorph with the long vowel will be selected in a non-diminutive plural NP if the plural marker is realised as *-en*.<sup>35</sup>

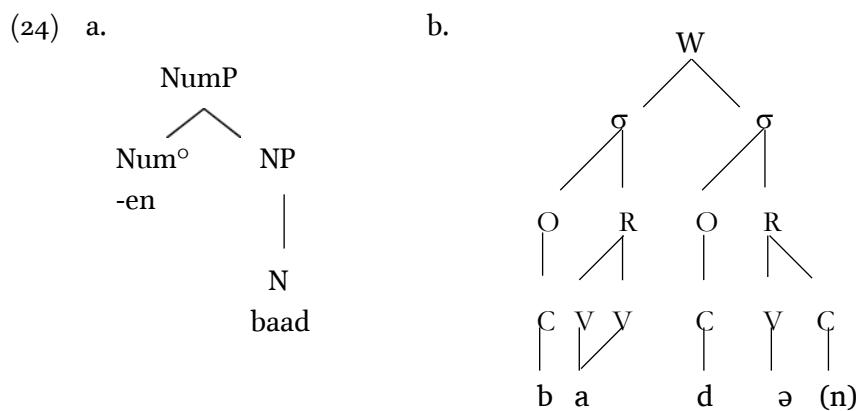
The data in this respect thus do not show some irregular behaviour that is specific to some vocabulary items. They allow for a generalisation. The fact that it is possible to generalise over data suggests that a rule is at work, rather than some stipulations at Vocabulary Insertion. Indeed, in what follows I discuss the fact that it is possible to account for the generalisation in (22) on the basis of the phonological rules in (15).

Van der Hulst (1985) pointed out that his approach to vowel length alternation presented in (15) captures the occurrence of the long vowel in non-diminutive plurals. When the plural marker *-en* attaches to a root the plural marker will occupy a separate syllable. Due to the Syllable Contact Law (Venneman 1988) which results in maximal onsets, the onset of this syllable contains the root's final consonant, as in (23).

Example (23) shows the plural of the noun *zoon* 'son'. Its stem has only one allomorph, which contains a long vowel. The example in (23)a represents the morphological structure,<sup>36</sup> (23)b shows the syllable structure.



If the root's rhyme consists of a short vowel followed by a single consonant the syllable structure of the plural form creates a context of ambisyllabicity. Hence, an allomorph with a long vowel is selected when available to avoid it, as proposed in (16).<sup>37</sup> The structure in (24)a is the morphological structure of *baad* 'baths', the one in (24)b is the phonological one.



<sup>35</sup> See (25) on why the plural marker *-en* is mentioned in this generalisation.

<sup>36</sup> For ease of exposition, I have represented the morphological structure in a conventional generative tree. It does not reflect my theoretical understanding of the noun phrase (see, for example, De Belder 2011a).

<sup>37</sup> Nothing hinges on the number of positions assigned to the schwa in the structure (see Trommelen 1984 and Van der Hulst 1985 for the debate on the number of moraic positions realised by the Dutch schwa).

The phonological approach immediately captures why the allomorph with the long vowel occurs if the root is string adjacent to the plural marker *-en*. In other words, it accounts for the omni-presence of the allomorph with the long vowel in the non-diminutive plural form. The stem allomorph with the short vowel, on the other hand, is selected when there are no phonological grounds to prefer a long vowel, as is the case in a singular NP, and it thus appears to be the elsewhere form.<sup>138</sup> Consider, in this respect, a vowel length alternating noun that may select both the plural marker *-en* and *-s*. As expected, the vowel will only lengthen in the phonologically relevant context, i.e. in front of *-en* (Booij 1998):<sup>139</sup>

- |   |                 |  |              |
|---|-----------------|--|--------------|
| (25) a. professoren<br>professor-PL<br>'professors' | [prɔfɛ'sorə(n)] | b. professor-s<br>professor-PL<br>'professors' | [prɔ'fɛsɔrs] |
|---|-----------------|--|--------------|

In the context of non-diminutive number inflection, the distribution of the stem allomorphs is thus fully and without any exception regulated by means of the phonological rules in (16).

### **8.3.3 The phonological pattern for verbs: Lengthening due to an abstract open syllable**

Zonneveld (1976, 1981-1982) points out that when stems for which an allomorph with a long vowel is available occur as weak verbs, the long vowel can be attested throughout the inflectional paradigm, even though the singular forms are not open syllables:

- |  |  |  |
|--|--|--|
| (26) ik baad<br>jij baadt<br>hij baadt<br>wij baden<br>jullie baden<br>zij baden | 'I bathe'<br>'you bathe'<br>'he bathes'<br>'we bathe'<br>'you bathe'<br>'they bathe' | [bat]<br>[bat]<br>[bat]<br>[badə(n)]<br>[badə(n)]<br>[badə(n)] |
|--|--|--|

Needless to say, the occurrence of the long vowel in the singular forms is unexpected based on their surface phonological forms; at least for the singular forms there does not seem to be a second syllable present that could trigger a problematic context of ambisyllabicity. At first sight, one might therefore guess that these patterns are irregular. However, there is no denying that the data behave in a strictly uniform pattern, which, indeed, allows for an empirical generalisation:

- (27) Morphologically simplex, weak verbs will select throughout the inflectional paradigm a stem allomorph with a long vowel, if available.

The issue has been addressed by Zonneveld (1976; 1981-1982). He captures these data by arguing that the underlying representations of these verbal forms do contain a second syllable, i.e. a schwa which

---

<sup>138</sup> The term *elsewhere form* is merely meant to describe the facts for now. In already indicated in Section 2.9 and as will be repeated in Section 6 it is actually not an elsewhere form from a technical point of view, but rather an unmarked form.

<sup>139</sup> A reviewer asked in which order Vocabulary Items are selected: do we first select the appropriate plural morpheme and then the appropriate stem allomorph or vice versa? The data in (26) suggest that the plural morpheme is selected first.

functions as the verb's theme vowel. Interestingly enough, he argues that the postulation of such a theme vowel does not only capture the presence of the long vowel, but also other, seemingly unrelated, unexpected phonological phenomena in the domain of Dutch verbs, such as d-weakening and schwa-deletion. For all these phenomena it is known independently that they occur in the presence of a vowel in Dutch. Taken together, these phenomena will allow the language acquiring child to assume the presence of such a theme vowel in the context of verbs (and only in the context of verbs!), even though it never surfaces. If one accepts Zonneveld's conclusions, the exception-less presence of the long vowel in the verbal forms in (26) follows immediately. For the plural forms the overt number inflection provides the required phonological environment, for the singular forms the underlying schwa provides the phonological context that triggers vowel length alternation, as summarized in (28).

- (28) Morphologically simplex, weak verbs will select throughout the inflectional paradigm a stem allomorph with a long vowel, if available, to satisfy the preference of vowel spreading over ambisyllabicity.

Note that for the singular forms the phonological allomorph selection takes into account the underlying representation.

### 8.3.4 Past tense plural forms

Zonneveld (1976; 1981-1982) noted that the past tense of some strong verbs in Dutch may show lengthening in the plural forms, i.e. those forms which receive an onset-less following syllable through number agreement. They show a short vowel in the singular.<sup>140</sup> Yet, a long vowel appears in the plural as the stem is followed by the plural inflectional marker *-ən*. Again, the data adhere to the same phonological principle:

- (29) Morphologically simplex, strong verbs with a plural agreement affix will select a stem allomorph with a long vowel to satisfy the preference of vowel spreading over ambisyllabicity.

This pattern is exception-less: all stems for which an allomorph with a long vowel is available are subject to the generalisation, which is illustrated in (30)-(32).<sup>141</sup>

(30)	a. ik at [at]	b. wij aat-en [atən]
	I ate	we ate-PL
	'I ate'	'we ate'

(31)	a. ik sprak [sprak]	b. wij spraak-en [sprakən]
	I spoke	we spoke-PL
	'I spoke'	'we spoke'

---

<sup>140</sup> It seems to follow that strong verbs do not select a theme vowel at the underlying representation.

<sup>141</sup> The exhaustive list of past tense stems for which a vowel length alternating allomorph is available, is: *at* 'ate', *beval* 'ordered', *bad* 'prayed', *brak* 'broke', *genas* 'healed', *gaf* 'gave', *kwam* 'came', *las* 'read', *lag* 'lied down', *mat* 'measured', *nam* 'took', *sprak* 'spoke', *stak* 'put', *stal* 'stole', *trad* 'went', *vergat* 'forgot', *vergaf* 'forgave', *vrat* 'hogged', *was* 'was', *zag* 'saw', *zat* 'sat'.

- (32) a. ik zag [zax]      b. wij zaag-en [zayən]  
           I saw                    we saw-PL  
           'I saw'                'we saw'

In short, the past tense of strong verbs shows the pattern which is expected on phonological grounds due to overt agreement affixes.

### 8.3.5 Conclusion

In the present section I have presented Van der Hulst's insight. Using a long vowel, when available, optimises syllable structure. It avoids ambisyllabicity in the presence of a following onset-less syllable. Three empirical domains are subject to this phonological condition, viz. plural non-diminutive nouns, Dutch morphologically simplex weak verbs and past tenses of strong verbs. For plural nouns, weak plural verbs and strong past plural verbs the following syllable is an overt inflectional morpheme, -ə(n). For singular weak verbs the following syllable is an abstract, underlying schwa.

Note that the principle is not only clearly optimising, but the same phonological principle has a predictable result across lexical items and across syntactic contexts. It should be clear that an approach that merely stipulates the occurrence of the stem allomorph as an insertion context at the level of the individual lexical items fails to capture the true nature of the phenomenon. The data rather suggest that the allomorph selection in this section is regulated by a phonological principle.

## 8.4 Irregular patterns

In the previous section we have seen domains for which a phonological principle determines the distribution of the stem allomorphs in an exception-less manner. The present section presents morphosyntactic domains for which this regularity breaks down. It will become clear that the phonological pattern still can be attested, yet it is subject to pervasive irregularities. These domains are diminutive forms and morphologically derived forms.

### 8.4.1 Diminutives

#### 8.4.1.1 The attested patterns of the diminutive forms

Dutch has one diminutive marker, viz. *-tje*.<sup>142</sup> It may surface as *-etje*, *-tje*, *-pje*, *-kje* or *-je* through phonological allomorphy (see Cohen 1958; Van Zonneveld 1978; Trommelen 1984; Van Oostendorp 1995; Van der Hulst 2008 and references therein).<sup>143</sup> Crucially, it generally does not give rise to any changes in the stem, as shown in (33).

- (33) a. een kat [kat]      b. een kat-je [kat<sup>j</sup>ə]  
           a cat                    a cat-DIM  
           'a cat'                'a small cat'

---

<sup>142</sup> Non-standard Dutch has the diminutive marker *-ke*.

<sup>143</sup> The choice of the diminutive allomorph depends on the final consonant and the stress pattern of the root it attaches to, see Section 4.1.2 for more details.

However, those nouns which do show a stem allomorph with a long vowel in the plural form, as presented in Section 3.2, may also show this stem allomorph in the diminutive form (see also Booij 1995: 72). The example is repeated in (34):



Interestingly, there are two patterns; the diminutive forms may or may not select the long vowel. These patterns are illustrated in Table 1, in which “S” refers to a stem with a short vowel and “L” refers to a stem with a long vowel:

**Table 1:** SL-patterns

	BASE FORM SG	BASE FORM PL	DIM SG	DIM PL
<b>SLSS</b> <i>bad</i> 'bath'	bad [ɑ]	baad-en [a]	bad-je [a]	bad-je-s [a]
<b>SLLL</b> <i>vat</i> 'barrel'	vat [ɑ]	vaat-en [a]	vaat-je [a]	vaat-je-s [a]

Nouns are distributed unevenly over these patterns. There are nine to ten SLLL nouns.<sup>144</sup> They are listed in (35):<sup>145</sup>

- |      |                 |            |              |         |               |          |            |          |
|------|-----------------|------------|--------------|---------|---------------|----------|------------|----------|
| (35) | <i>blad</i>     | 'leaf'     | <i>glas</i>  | 'glass' | <i>pad</i>    | 'path'   | <i>vat</i> | 'barrel' |
|      | <i>demon</i>    | 'demon'    | <i>gat</i>   | 'hole'  | <i>proton</i> | 'proton' |            |          |
|      | <i>elektron</i> | 'electron' | % <i>lot</i> | 'lot'   | <i>schip</i>  | 'ship'   |            |          |

All other vowel length alternating nouns are SLSS nouns. They clearly constitute the largest group:<sup>146</sup>

<sup>144</sup> The noun *lot* 'lot' is an SLLL noun in Northern Dutch. It is an SLSS noun in Belgian Dutch.

<sup>145</sup> In (56) it is pointed out that *proton* ‘proton’, *elektron* ‘electron’ and *demon* ‘demon’ contain a vowel alternating suffix *-on*. As such, all nouns containing this suffix will show this pattern. The claim that there are ten nouns is strictly speaking thus inaccurate. A reviewer further mentions nouns ending in *-ol*: *fenolen* ‘phenols’, *alcoholen* ‘alcohols’. I assume *-ol* is a vowel alternating suffix as well.

<sup>146</sup> There is a noun, *satan* ‘satan’, which arguably belongs to the SLSS group as it has a short vowel in the diminutive forms. However, it selects the plural affix *-s* (*satans*) and therefore strictly speaking does not belong to this group, as its non-diminutive plural form does not select the stem with the long vowel either. As such, it is phonologically well-behaved. We do see the long vowel appear in derived forms, such as *satanisch* ‘satanic’.

(36)	<i>bad</i>	'bath'	<i>gebrek</i>	'deficiency'	<i>lid</i>	'member'	<i>staf</i>	'rod'
	<i>bedrag</i>	'amount'	<i>gen</i>	'gene'	<i>literator</i>	'writer'	<i>smid</i>	'smith'
	<i>dag</i>	'day'	<i>god</i>	'god'	<i>% lot</i>	'lot'	<i>stad</i>	'city'
	<i>dak</i>	'roof'	<i>graf</i>	'grave'	<i>oorlog</i>	'war'	<i>tred</i>	'pace'
	<i>dal</i>	'valley'	<i>hertog</i>	'duke'	<i>pastor</i>	'priest'	<i>verbod</i>	'ban'
	<i>gebed</i>	'prayer'	<i>hof</i>	'court'	<i>professor</i>	'professor'	<i>verdrag</i>	'treaty'
	<i>gebod</i>	'command'	<i>hol</i>	'hole'	<i>slag</i>	'stroke'	<i>weg</i>	'road'

Nouns thus fall in two different groups. For good measure, note that these groups and the nouns they contain only represent the judgments of the majority of speakers. The facts are certainly subject to idiolectal variation.

The SLSS group is not only the largest group, it also seems to be the more regular group. A reviewer pointed out that if a neologism acquires a stem allomorph with a long vowel, it seems to belong to this pattern, as shown in (37).<sup>147</sup>

(37) a.	<i>een gen</i>	[yɛn]	b.	<i>twee geen-en</i>	[yɛnə(n)]
	a gene			two gene-PL	
	'a gene'			'two genes'	
c.	<i>een gen-etje</i>	[yɛnət̪ə]	d.	<i>twee gen-etje-s</i>	[yɛnət̪əs]
	a gene-DIM			two gene-DIM-PL	
	'a small gene'			'two small genes'	

Our task is thus twofold. Firstly, we have to account for the fact that there are two patterns, viz. SLLL and SLSS. Secondly, we have to capture the fact that the second pattern, viz. SLSS, seems to be the more general pattern. In what follows I will discuss which patterns we would expect if phonology alone determined the distribution of the allomorphs in this domain. It will become clear that the empirical domain is hybrid: the majority of the cases can be captured on phonological grounds, but some diminutive forms show exceptional behavior and must simply have been lexicalised.

#### 8.4.1.2 *The diminutive affix*

We have seen in Section 3 that stems with long vowels are selected to avoid ambisyllabicity. Ambisyllabicity occurs if the root contains a short vowel and its final consonant becomes the onset of the following syllable due to the Syllable Contact Law. One thus predicts that if the root's final consonant(s) can become the onset of the following syllable, the allomorph with the long vowel will be selected and vice versa. In order to understand the consequences of these facts for the diminutive forms, let us check how a diminutive form is syllabified.

Recall that the Standard Dutch diminutive affix has several allomorphs, viz. *-etje*, *-etje*, *-je*, *-pje* and *-kje* ([ət̪ə], [t̪ə], [jə], [pjə], [kjə]). Their distribution has been studied before (see Cohen 1958; Trommelen 1984; Van der Hulst 2008 and references therein) and the facts are rather intricate. For the data under discussion it is important to note that a short vowel followed by an obstruent will select *-je*, a stressed short vowel followed by a sonorant selects *-etje* and a non-stressed short vowel

<sup>147</sup> I agree with the judgments of the reviewer presented in (39). Another reviewer, however, finds the forms with the long vowel acceptable as well. The diminutive forms of the word are too infrequent to occur in corpora. I conclude that the claim in (39) is uncertain.

or a long vowel followed by a sonorant selects *-tje*. The [t] in *-tje* is probably not a /t/ at the underlying representation, but rather a stop unspecified for place (Van der Hulst 1984), its default value being coronal.<sup>148</sup> As such, it will assimilate for place with a stem's final nasal and surface as [t̪ə], [pjə] or [kjə]. In short, it is clear that the root's phonological make-up will have an influence on the form of the diminutive affix. I therefore discuss roots with different final segments in different subsections.

Diminutives formed with the allomorph *-etje* show exceptional phonological behavior in various respects (Lowenstamm and van der Wilt 1982; Van der Hulst 2008). A discussion of these long forms would take us too far afield, even though the list of vowel length alternating roots does contain roots which do select *-etje*, viz. *dal* 'valley' and *hol* 'hole'. I will omit them from the discussion and I will restrict the present discussion to the other listed roots.

#### *8.4.1.3 Stems ending in an obstruent other than /t/, /d/ or /p/*

As I will describe in the following sections, there are reasons to consider the phonology of the diminutive of roots ending in a /t/, /d/ or /p/ separately. The current section discusses those vowel alternating stems that end in an obstruent other than /t/, /d/ or /p/. We will see that we expect the short vowel to occur and, indeed, the short vowel is most often attested in these diminutive forms, yet not without an exception.

As mentioned above, stems with a short vowel ending in an obstruent other than /t/, /d/ or /p/ select the diminutive affix *-je*:

- |                        |                   |
|------------------------|-------------------|
| (38) a. dak-je [dakjə] | b. dag-je [daxjə] |
| roof-DIM               | day-DIM           |
| 'small roof'           | 'short day'       |

Hypothetically, the final consonant of the stem could occupy the following positions in the syllabic structure: it could realise the coda of the final stem syllable, it could realise the onset of the following one or it could realise both, in which case it would be ambisyllabic. The data suggest that it only realises the coda of the final stem syllable. The first indication is that an underlying voiced final obstruent, which surfaces in the plural forms as shown in (39)a surfaces as a devoiced obstruent in the diminutive forms, as shown in (39)b:

- |                          |                   |
|--------------------------|-------------------|
| (39) a. hov-en [hovə(n)] | b. hof-je [hɔfjə] |
| garden/court-PL          | garden/court-DIM  |
| 'gardens/courts'         | 'gardens/courts'  |

It is well-known that Dutch has final devoicing at syllable boundaries (Booij 1995; Kooij & Van Oostendorp 2003). We therefore must conclude that the obstruents occupy (at least) the coda of the final stem syllable, where it would occupy a syllable boundary. An indication that the obstruent does not occupy the onset of the following syllable as well, is the fact that C<sub>j</sub>/C<sup>j</sup>-clusters do not seem to exist in Dutch (cf. Van Oostendorp 1995 for a careful and much more detailed discussion).<sup>149</sup>

---

<sup>148</sup> Van Oostendorp (1995) proposes it is an empty consonantal root, which is also the representation of a glottal stop in his analysis.

<sup>149</sup> The facts are much more complex as the present section suggests, as the underlying form of the diminutive is arguably /iə/ (see Van Oostendorp 1995). As a result, the /i/ present in the underlying representation can surface both as a vowel,

Admittedly, there is the occasional loanword which shows a C<sup>j</sup>/C<sup>j</sup>-cluster as an onset, both in Standard Dutch, as in (40), and in dialectal words, as in (41).

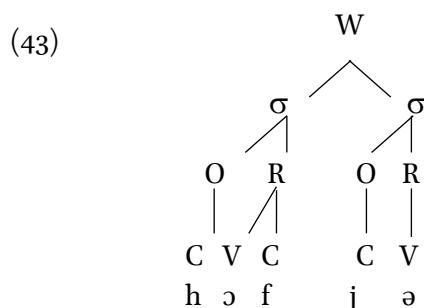
- (40) a. garage [ga'raʒə]      b. fjord [fjɔrt]  
       'garage'                          'fjord'  
       (from French *garage*)         (from Norwegian *fjord*)
- c. sjaal [ʃal]  
       'shawl'  
       (from English *shawl*, French *châle*)

- (41) a. anzjoen<sup>150</sup> [an'ʒun]      b. nondedju [nɔndə'dʒy]  
       'onion'                          'goddamn'  
       (from French *oignon*)         (from French *nom de Dieu*)

A C<sup>j</sup>-onset may equally occur in onomatopoeia's, as in (42).<sup>151</sup>

- (42) a. d(z)joeven [dʒuvən]      b. % djakken [dʒakən]  
       'to whiz'                        'to swish'

Yet, apart from these exceptions, it does not seem to be part of Dutch phonology. I therefore assume that the onset of the diminutive affix syllable is solely realised by the glide contained in the diminutive affix, as represented in (43).



Note that the first syllable in the structure is a closed, bimoraic syllable and, as such, it is a well-formed Dutch syllable. Furthermore, there is no ambisyllabicity in the structure that is to be avoided. In short, the stem with the short vowel already fulfills the requirements of the syllable structure and there is no reason to select the stem allomorph with the long vowel. On phonological grounds, we thus expect the stem with the short vowel for all stems ending in an obstruent other than /t/, /d/ or /p/. In other words, for such stems, we expect nothing but SLSS nouns.

Let us now compare these expectations with the data. Indeed, most nouns of this format are SLSS nouns:

---

realising the nucleus of the diminutive, and as a glide, realising the onset of the diminutive affix, in Dutch dialects. However, this complication is not relevant for the present discussion and I am therefore abstracting away from this issue for ease of exposition.

<sup>150</sup> See [www.vlaamswoordenboek.be](http://www.vlaamswoordenboek.be)

<sup>151</sup> The examples are restricted to non-standard Belgian Dutch.

(44)	<i>bedrag</i>	'amount'	<i>graf</i>	'grave'	<i>slag</i>	'stroke'
	<i>dag</i>	'day'	<i>hertog</i>	'duke'	<i>staf</i>	'rod'
	<i>dak</i>	'roof'	<i>hof</i>	'garden'	<i>verdrag</i>	'treaty'
	<i>gebrek</i>	'deficiency'	<i>oorlog</i>	'war'	<i>weg</i>	'road'

Yet, within this group, we also find an SLLL noun:

- (45)    *glas*    'glass'

The observation is remarkable. The group as a whole shows an expected pattern on phonological grounds and we understand why the majority of nouns selects which allomorph. Yet, undeniably, there is an exception. As discussed at length in Section 2 the field goes back and forth on focusing on the phonologically expected selection of allomorphs and on emphasising that phonology, in fact, fails to capture the distribution of allomorphs. The data, at this point, however suggest that we should acknowledge that both positions are equally valid: phonology does seem to influence the distribution of allomorphs, while clear exceptions exist as well. I argue that this observation is not problematic, but indicative of the way allomorph selection is organised in the Y-model.

#### 8.4.1.4 Stems ending in a /d/, a /t/ or a /p/

In the previous section, we concluded that a final obstruent of a stem occupies solely the coda of the final stem syllable in diminutive forms. It is not possible to immediately adopt this insight for stems ending in a /d/ or a /t/. Firstly, notice that the coronal stops palatalise in diminutive forms:

- (46) a. *bad-je*                         [bat<sup>j</sup>ə]  
             bath-PL  
             'baths'  
         b. *lot-je*                         [lɔt<sup>j</sup>ə]  
             lot-DIM  
             'lottery ticket' (Belgian Dutch)

In a discussion of the Dutch clitic *je* /iə/, which may follow verbal stems ending in a /d/ or a /t/, Van Oostendorp (1995: 259-261) proposes that the /i/ and the stem's final coronal stop merge into a single segment, which is realised as a palatalised coronal onset. The proposal can be adopted here as it accounts for the fact that the coronal and the glide are indeed realised as a single segment. We have to conclude that the coronal thus occupies at least the onset of the diminutive affix syllable.

Now, does it simultaneously realise the coda of the final stem syllable in an ambisyllabic setting? Well, it is clear that an underlying /d/ is devoiced in the diminutive form, as can be seen in (48a). However, for these cases the devoicing does not necessarily result from final devoicing, which would indicate the presence of a syllable boundary. Van Oostendorp (1995: 266) argues that the mere process of palatalisation itself already causes the devoicing of the /d/. It is therefore hard to determine whether the /d/ and the /t/ occupy the coda of the final stem syllable as well in the examples in (48).

However, we do not even have be sure about whether /d/ and /t/ occupy the coda of the final stem syllable to formulate the phonological expectations. The mere fact that it does occupy the onset of the diminutive affix syllable suffices to expect the stem with the long vowel, either because the previous syllable ends up being an open syllable or because its final consonant ends up being ambisyllabic. Recall that we have seen in (16) that the selection of an allomorph with a long vowel trumps an ambisyllabic configuration. Given that for these stems such an allomorph is available,

ambisyllabicity should be avoided. We thus expect the SLLL pattern on phonological grounds. Interestingly enough, this is not exactly what we attest: both types exist. The SLSS nouns ending in a /d/ or a /t/ are given in (47) and (48):

- (47) *bad* 'bath'      *god* 'god'      *stad* 'city'  
*gebed* 'prayer'      *lid* 'member'      *tred* 'pace'  
*gebod* 'command'      *smid* 'smith'      *verbod* 'ban'

- (48) *lot* 'lot' (Belgian Dutch)

The SLLL nouns ending in a /d/ or a /t/ are given in (49) and (50):

- (49) *blad* 'leaf'      *pad* 'path'

- (50) *gat* 'hole'      *vat* 'barrel'      *lot* 'lot' (Northern Dutch)

For the stems ending in a /d/, the SLSS pattern unexpectedly outnumbers the phonologically expected SLLL pattern. One might either take this as an indication that Van Oostendorp's analysis is on the wrong track. Alternatively, one has to conclude that speakers of Dutch are fully capable of storing nine nouns with exceptional behaviour in their lexicon. I adopt the second alternative.

On a final note, there is a single noun, *schip* 'ship' of which the stem ends in a /p/ and which is an SLLL noun. I have to admit that I am not even sure which pattern we expect for this noun on the basis of Dutch phonology. We have seen that Cj-clusters are not attested in the phonology of Dutch. Yet, the clusters [p<sup>j</sup>] and [t<sup>j</sup>] clearly occupy onsets when they are the onset of a diminutive allomorph. This becomes clear through the fact that the allomorphs *-tje* and *-pje* may attach to roots which have maximal rhymes.<sup>152</sup> Dutch rhymes contain at most three positions (Trommelen 1984):

- (51) a. ✓ *zoon* 'son'      [zon]  
b. ✓ *lamp* 'lamp'      [lamp]  
c. \* *laamp*      [lamp]

As such, one can deduce the fact that the syllable structure of the following diminutives is aligned with their morphological structure:

- (52) a. *zoon-tje*      [zon<sub>σ</sub>-t<sup>j</sup>ə<sub>σ</sub>]  
son-DIM  
'little son'  
b. *bloem-pje*      [blum<sub>σ</sub>-pjə<sub>σ</sub>]  
flower-DIM  
'small flower'

After all, if the initial consonant of the diminutive affix were part of the final stem syllable, this syllable would contain four positions, which is illicit. As such, we can conclude that [pj] and [t<sup>j</sup>] do occur as onsets in Dutch, at least for diminutives. As we have seen above, this is fully expected for the [t<sup>j</sup>] cluster due to the palatalisation of the coronal stop. However, it is a surprising observation

---

<sup>152</sup> The diminutive allomorph *-kje* does not attach to maximal rhymes.

for the [pj]-cluster. Outside of the domain of the diminutive, it is not attested as an onset and, at least in Standard Dutch, the /p/ does not palatalise. Yet, the example in (54b) clearly suggests the /p/ may occupy an onset position. As such, one might conjecture that the SLLL pattern of *schip* is the phonologically expected pattern. To the best of my knowledge, the fact in (52) has not been observed, let alone captured, before and I have nothing to add beyond the observation.

I can add to the observation that precisely those stops that show the pattern illustrated in (52), also disproportionately occur as the final segment of SLLL nouns. (53) shows all vowel alternating SLSS nouns that end in an obstruent, (54) shows all the SLLL ones.

- |      |                            |                       |               |              |              |               |             |                |          |
|------|----------------------------|-----------------------|---------------|--------------|--------------|---------------|-------------|----------------|----------|
| (53) | <i>bad</i>                 | 'bath'                | <i>gebrek</i> | 'deficiency' | % <i>lot</i> | 'lot'         | <i>tred</i> | 'pace'         |          |
|      | <i>bedrag</i>              | 'amount'              | <i>god</i>    | 'god'        |              | <i>oorlog</i> | 'war'       | <i>verbod</i>  | 'ban'    |
|      | <i>dag</i>                 | 'day'                 | <i>graf</i>   | 'grave'      |              | <i>smid</i>   | 'smith'     | <i>verdrag</i> | 'treaty' |
|      | <i>dak</i>                 | 'roof'                | <i>hertog</i> | 'duke'       |              | <i>slag</i>   | 'stroke'    | <i>weg</i>     | 'road'   |
|      | <i>gebed</i>               | 'prayer'              | <i>hof</i>    | 'garden'     |              | <i>stad</i>   | 'city'      |                |          |
|      | <i>gebod</i>               | 'command'             | <i>lid</i>    | 'member'     |              | <i>staf</i>   | 'rod'       |                |          |
| (54) | <i>schip</i>               | 'ship'                |               |              | <i>vat</i>   | 'barrel'      | <i>blad</i> | 'leaf'         |          |
|      | <i>gat</i>                 | 'hole'                |               |              | <i>glas</i>  | 'glass'       | <i>pad</i>  | 'path'         |          |
|      | % <i>lot</i> <sup>53</sup> | 'fate/lottery ticket' |               |              |              |               |             |                |          |

A Fisher's exact test - a test which is designed to test data sets with a limited number of attestations - suggests that this pattern is significant ( $p < 0.05$ ).<sup>54</sup> At this point, I fail to understand why this should be the case.

In this section we have seen varied behaviour. Nouns ending in a /d/ are predominantly SLSS nouns, whereas we expect them to be SLSS nouns. Nouns ending in a /t/ are expected to be SLLL nouns as well and they mostly seem to adhere to this pattern. Finally, the noun *schip* 'ship' is an SLLL noun as well and it is hard to tell whether this is expected as there is something about the behaviour of the [p] in the diminutive affix that we fail to understand.

It becomes clear that the diminutive forms readily allow for exceptions, which have to be listed as such in the speaker's lexicon. One may wonder if not simply all diminutive forms of vowel alternating stems are stored in the speaker's lexicon. Admittedly, on the basis of the data, it is impossible to tell. For mere reasons of parsimony, I will continue to assume that what can be derived on phonological grounds is derived as such: there is no need to store forms that are derivable.

#### 8.4.1.5 Stems ending in a sonorant

I have discussed the syllable structure of diminutives of which the roots end in an obstruent. For good measure, I will briefly comment on roots ending in a sonorant with a non-stressed final syllable.<sup>55</sup> Such roots select the affix *-tje*:

---

<sup>53</sup> The noun *lot* 'lot' is only an SLLL noun in the Netherlands.

<sup>54</sup> It does not matter in which group one counts *lot*, the result is significant in both scenarios. I would like to thank a reviewer for suggesting this test.

<sup>55</sup> Roots ending in a sonorant with a stressed final syllable select the so-called long form of the diminutive, viz. *-etje*. I ignore these forms as they would take us too far afield (see van der Hulst 2008 and references therein for discussion).

- (55) literator-tje [lite'rator]  
writer-DIM  
'small writer'

There are no clusters such as [rt<sup>j</sup>] or [nt<sup>j</sup>] which function as onsets in Dutch. The /r/ therefore occupies the coda of the final stem syllable and there is no need to use the root's allomorph with the long vowel: we expect the SLSS ending. This is indeed what we attest for nouns such as *literator* 'literary author', *professor* 'professor' and *pastor* 'priest'.

A reviewer points out that several nouns ending in *-on* actually show the SLLL pattern:<sup>156</sup>

- (56) demon 'demon' [ɔ] demoon-tje 'small demon' [o]  
elektron 'elektron' [ɔ] elektroon-tje 'small electron' [o]  
proton 'proton' [ɔ] protoon-tje 'small proton' [o]

I suggest that for these cases we are actually systematically dealing with one and the same suffix *-on*, which is recognisable as a suffix to the language learner by the fact that it has the following properties: it attaches to classic roots and it attracts stress in its non-diminutive singular form, whereas nouns ending in *-on* otherwise typically bear stress on the final syllable:

- (57) 'demon 'demon'  
e'lektron 'elektron'  
'proton 'proton'
- (58) ja'pon 'gown'  
kar'ton 'cardboard'  
be'ton 'concrete'

Rather than several exceptions, this class of nouns thus instantiates only a single exception, which happens to be a suffix. In sum, the phonologically expected pattern here is the SLSS noun, which is indeed attested in the majority of the cases.

#### 8.4.1.6 Lexicalisation

Let us summarise what we have observed so far for diminutive forms. There is a phonologically expected pattern and, often, the majority of the nouns adhere to this pattern. More specifically, we understand why SLSS nouns in general outnumber the SLLL nouns. Yet, simultaneously, there are systematic exceptions. There is even a noun, *lot* 'lot', which is subject to regional variation. The group as a whole simply cannot be accounted for on purely phonological grounds. Note that the exceptions are even clearly non-optimal from a phonological point of view. As such, they echo Paster's (2006) observation that allomorph selection may undermine phonological requirements. She concluded that allomorph selection should be listed. For at least the exceptional examples we

---

<sup>156</sup> The diminutive forms of these nouns are actually not listed in any descriptive or prescriptive sources and they are not common in corpora such as *Neder-Lab* either. Yet, I agree with the native speaker's judgments of the reviewer. See further footnote 16 on the suffix *-ol*.

have encountered, I concur with Paster that stipulating the exceptional behaviour at Vocabulary Insertion is indeed what should be done.

There is thus another selection mechanism for allomorphy, apart from phonology, viz. listing. More generally, it is clear that the selection of an allomorph may simply be a lexically stored irregularity, also outside of the domain of vowel length alternation. We have seen that the participle of the verb *uitverkiezen* ‘to select’ is *uitverkozen*, as can be seen in (59):

- (59) Jij bent uitverkozen.  
you are selected  
'You are selected.'

Yet, there is an idiomatic reading in which a different form is used:

- (60) het uitverkoren volk  
the selected people  
'the Chosen People (i.e. the Jews)'

There is nothing systematic to be mentioned about the exceptional use of *uitverkoren* in the example above. A connection has been stored between an irregularity in form and an irregularity in meaning in a specific collocation. The use of an irregular form has simply been listed as the conventionalised form to express a specific idiomatic meaning. The selection of the allomorph is then an arbitrary convention, a conventionalized irregularity. An allomorph may thus be selected in a specific context despite the fact that there are no phonological or syntactic grounds to do so. I propose this is what is happening for the nouns which are not accounted for on the basis of phonology. The data in the following sections further support this view.

#### 8.4.2 Derivations of vowel alternating stems

So far we have seen that many occurrences of the stem allomorph with the long vowel can be ascribed to phonology: if a following syllable causes the stem to be an open syllable, an allomorph with a long vowel will be selected, if available, to avoid ambisyllabicity. We have seen two domains in which this pattern is exception-less, viz. nominal plural marking and verbs. In contrast, we have also encountered a domain in which this phonological rule seemed to lead to a strong tendency rather than an empirical generalisation: the distribution of stem allomorphy for diminutive forms largely adhered to phonological generalisations, but not in an exception-less manner. Lexicalised exceptions seemed to play a role in this domain. In the present section I discuss data from the domain of morphological derivation. They behave on a par with the diminutives: most often one attests the allomorphic distribution which is expected from a phonological point of view, yet not without exception.

In Section 3.1 I already cited the following paradigm noted by Van der Hulst (1985: 63), which shows stems merging with a derivational affix:

- |      |    |        |     |          |            |     |             |
|------|----|--------|-----|----------|------------|-----|-------------|
| (61) | a. | demon  | [ɔ] | 'demon'  | demonisch  | [o] | 'demonic'   |
|      | b. | satan  | [a] | 'satan'  | satanisch  | [a] | 'satanic'   |
|      | c. | Israel | [ɛ] | 'Israel' | Israelisch | [e] | 'Israeliic' |

I discussed that the long vowel appears in context that would otherwise cause ambisyllabicity. Similarly, in (62)a-b, vowel length alternating stems merge with the derivational affix *-elijk*, causing the stem to become an open syllable. Again, the stem with the long vowel is expected on phonological grounds. The examples contrast with the example in (62)c, a noun for which no stem allomorph is available:<sup>157</sup>

- (62) a. dag [a] 'day' dagelijks [a] 'daily'  
 b. hertog [ɔ] 'duke' hertogelijk [o] 'ducal'  
 c. bisschop [ɔ] 'bishop' bisschoppelijk [ɔ] 'episcopal'

In short, we see again that, quite generally, if a stem occurs in a phonological environment that causes it to appear as an open syllable and if a stem allomorph with a long vowel is available, that stem allomorph will be selected in Dutch to avoid ambisyllabicity.

Yet, a few counterexamples are attested. Booij (1999: 88) mentions the following examples:

- (63) a. schip-er [sxipər]  
 ship-SUFFIX<sub>N</sub>  
 'skipper'  
 b. spel-etje [spɛlətjə]  
 game-DIM  
 'game'  
 c. god-elijk [yɔdəlk]  
 god-SUFFIX<sub>N</sub>  
 'divine'  
 d. bad-er-en [badərə(n)]  
 bath-SUFFIX<sub>V-INF</sub>  
 'to bathe' (informal)

Another example worth mentioning in this respect is the following:

- (64) gebrek-ig [yəbrækəx]  
 deficiency-SUFFIX<sub>A</sub>  
 'defective'

In all examples, the stem contains a single final consonant and it is followed by a vowel, making the stem an open syllable. Yet, an allomorph with a long vowel does not appear. In sum, derivational morphology shows instances in which allomorph selection is but a stored irregularity. As was the case for diminutives, these instances are clearly the exceptions and do not form the regular pattern. However, there is no denying they exist.

So far, I have discussed examples in which the allomorph with the short vowel appears while the one with the long vowel is expected phonologically. The reverse situation occurs as well. Allomorphs with long vowels may be attested in adjectives and adverbs that do not show the expected phonological structure at all; the stem is not an open syllable. Examples are shown in (65).

- (65) a. van-daag [vəndax]  
 of-day  
 'today'  
 b. daag-s [daxs]  
 day-SUFFIX<sub>Adv</sub>  
 'daily'

---

<sup>157</sup> I would like to thank an anonymous reviewer for mentioning these data.

The allomorph invariably stands in an environment in which the allomorph is selected on phonological grounds in earlier stages of Dutch. It is a reasonable assumption that, subsequently, they have been lexicalised as such. This applies to the forms in (65). They are both derived from forms which were inflected for case and as such contained an extra syllable (see *Woordenboek der Nederlandsche Taal*):

- |  |   |
|--|---|
| (66) <i>Modern Dutch</i>                           | <i>Middle Dutch</i>                       |
| a. van-daag<br>of-day<br>'today'                   | b. van-daag-e<br>of-day-DATIVE<br>'today' |
|  |   |
| (67) <i>Modern Dutch</i>                           | <i>Middle Dutch</i>                       |
| a. daag-s<br>day- SUFFIX <sub>Adv</sub><br>'daily' | b. daag-es<br>day-GENITIVE<br>'daily'     |

The occurrence of the allomorph in these environments further supports the claim that the distribution of the vowel alternating allomorphs may rely on lexicalisation.

## 8.5 Illicit patterns

I have discussed allomorphy in various domains: nouns, verbs and morphological derivations. We have seen on the basis of attested data that two factors are at play regulating the distribution of vowel alternating allomorphs in Dutch: phonology and lexicalisation. In this section I briefly come back to the nouns and I discuss what is not attested. They illustrate again that phonology is indeed at play in these matters. Furthermore, they indicate that syntactic insertion contexts or syntactic features are not relevant for this empirical domain in Dutch.

### 8.5.1 Principally excluded patterns

We have seen that vowel length alternation in Dutch shows two patterns, viz. SLSS and SLLL, as summarised in Table 1, which is repeated in this section (S: stem with a short vowel, L: stem with a long vowel).

**Table 1 (repeated): SL-patterns**

	BASE FORM SG	BASE FORM PL	DIM SG	DIM PL
<b>SLSS</b> <i>bad</i> 'bath'	bad /a/	baad-en /a/	bad-je /a/	bad-je-s /a/
<b>SLLL</b> <i>vat</i> 'barrel'	vat /a/	vaat-en /a/	vaat-je /a/	vaat-je-s /a/

It is striking that only these patterns exist. They certainly do not exhaust all hypothetical possibilities as listed in Table 2. The present section addresses this issue.

**Table 2:** exhaustive overview of SL-patterns

	BASE FORM SG	BASE FORM PL	DIM SG	DIM PL
*SLLS	short	long	long	short
*LSSS	long	short	short	short
*LSLS	long	short	long	short
*LSSL	long	short	short	long
*LSLL	long	short	long	long
*SLSL	short	long	short	long
SSSS	short	short	short	short
LLLL	long	long	long	long

Let me start with a remark. One could argue that, in principle, any pattern could be stored as a lexicalised pattern and according to this logic no patterns should be excluded. Even though this is a fair point in principle, it is not a reasonable expectation. Lexicalised exceptions historically go back to forms that once were motivated. More specifically, they go back to forms that did contain the relevant phonological context to trigger the allomorph in the past, as we have seen in Section 4.3 (and see Lahiri & Dresher 1999 on Open Syllable Lengthening in West-Germanic). As a result, we do find lexicalised exceptions in narrowly defined domains, such as the diminutive or morphological derivations and they are systematically absent from other domains.

First consider the following hypothetical patterns: SSSL, SSLL and SSLS.<sup>158</sup> These patterns express that a hypothetical noun, for which a stem allomorph with a long vowel exists -as shown by a diminutive form- fails to select this allomorph for the non-diminutive plural. In other words, it fails to grasp the opportunity to optimise the non-diminutive plural phonologically. Such nouns are not attested.<sup>159</sup> Note that this observation strongly supports the claim that phonological optimisation is the main factor in the distribution of vowel alternating stem allomorphs in Dutch.

The SLLS pattern is phonologically inconsistent as both for the singular and the plural diminutive form the root is followed by the exact same diminutive affix. Hence, given that they share the same phonological context, it is predicted they should contain the same allomorph.

The illicit patterns LSSS, LSLS, LSSL, LSLL express that a vowel shortens in the regular plural. Such a process would be far from optimal by the principles of Dutch syllabification as it creates the dis-preferred ambisyllabicity in the non-diminutive plural form.<sup>160</sup> There is yet another aspect that

<sup>158</sup> Recall from Section 2 that Dutch has two affixes to mark plurality, viz. *-en* and *-s*. As such, the SSLL pattern could have existed if there had been a noun that selected the affix *-s* in the regular plural form and the long form in the diminutive forms.

<sup>159</sup> As noted above, there is the noun *satan* ‘satan’, which shows the short vowel in the plural form. Yet, this is expected, as it selects the plural affix *-s*: *satans*.

<sup>160</sup> A reviewer points out that s/he agrees that there is indeed no reason to assume vowel shortening for the Standard Dutch syllable. However, it should be kept in mind that closely related variants, which arguably have different syllabic

we have not addressed so far that renders these forms far from optimal. Kager and Zonneveld (1985–1986) argue that the Dutch rhyme is bimoraic and that the final consonant of a trimoraic syllable occupies a position in the syllable's appendix. Empirical support for this claim comes from the fact that trimoraic syllables are restricted to word-final positions. Given that appendices by definition only occur word-finally, the observation is immediately captured by proposing that the Dutch rhyme is in fact bimoraic. It is therefore reasonable to assume that superheavy syllables are less optimal than heavy syllables. It now immediately follows why a long vowel will not appear in the non-diminutive singular. A short vowel results in a bimoraic rhyme, a long vowel would result in a more marked syllable, which has to assign one consonant to the appendix. We thus consequently expect the short vowel in the non-diminutive singular and this is exactly what we see. This observation is not trivial. Again, it indicates that phonological principles are an important factor in the distribution of allomorphs in this domain.

The fact that we have phonological reasons to assume the short vowel in the non-diminutive singular form implies that this form is not *per se* a *default* form at Vocabulary Insertion. It seems to be empirically incorrect to state that it occurs in elsewhere positions. It rather seems to occur in those positions where its moraic format is actually preferable.

Given this state of affairs, one can only choose for a long vowel in the non-diminutive singular if there is no allomorph with a short vowel available at all, i.e. LLLL. Note, trivially, that the LLLL pattern exists. Dutch has many stems with a long vowel for which no allomorph with a short vowel is available. Similarly, the SSSS pattern exists as well; Dutch has many stems with a short vowel for which no allomorph with a long vowel is available.

In sum, we actually understand on phonological grounds not only why some patterns do occur, but also why many patterns do not occur. These observations again strongly support the claim that Phonology does determine the selection of the allomorph for many of these cases. By now, we have a lot of descriptive adequacy to lose by assuming that all allomorph selection in Dutch simply happens to be listed. It is much more advantageous to propose that Phonology determines the allomorph distribution for the majority of cases, despite the fact that for some forms we have to assume listedness.

### 8.5.2 Accidentally excluded patterns: syntactic patterns

There is one pattern I have not discussed so far, viz. the SSL pattern. The fact that this pattern is absent is not a theoretical necessity, but an accidental gap. As discussed earlier in Section 2.3 syntactic plurality in itself may select allomorphs or other irregular exponents. Yiddish, for example, shows allomorphy in plural contexts. In other words, it has an XYXY pattern (Lowenstamm 2007). In the following example the root *kind* 'child' alternates with its allomorph *kinder* in an XYXY fashion:

- |      |                             |                                     |                                      |   |
|------|-----------------------------|-------------------------------------|--------------------------------------|---|
| (68) | a. kind<br>child<br>'child' | b. kinder<br>children<br>'children' | c. kind-l<br>child-DIM<br>'children' | d. kinder-l-ex<br>child-DIM-PL<br>'little children' |
|------|-----------------------------|-------------------------------------|--------------------------------------|---|

---

requirements, do show vowel shortening. Relevant examples are Frisian (Tiersma 1985, Postma 1990), Low German (Stiebels 2013) and West-Flemish (as observed by the reviewer). The absence of vowel shortening in Dutch thus cannot simply be a universal (or West Germanic) tendency, but must be related to its syllabic structure.

In short, an irregular exponent may be found in syntactically plural contexts. Similar data are found in Amharic (Kramer 2009).

Vowel length alternating nouns are not the only nouns that show stem allomorphy in Dutch. Just as in Yiddish, Dutch has stems that alternate between a simple form and an *-er* form.<sup>161</sup> As in Yiddish, these nouns actually show an XYXY pattern:

- |         |         |    |                         |    |                |    |                             |      |                          |
|---------|---------|----|-------------------------|----|----------------|----|-----------------------------|------|--------------------------|
| (69) a. | kind    | b. | kinder-en               | c. | kind-je        | d. | kinder-tje-s                | e. * | kinder-tje               |
|         | child   |    | child <sub>ALL-PL</sub> |    | child-DIM      |    | child <sub>ALL-DIM-PL</sub> |      | child <sub>ALL-DIM</sub> |
|         | 'child' |    | 'children'              |    | 'little child' |    | 'little children'           |      |                          |
- 
- |         |       |    |                       |    |             |    |                           |      |                        |
|---------|-------|----|-----------------------|----|-------------|----|---------------------------|------|------------------------|
| (70) a. | ei    | b. | eier-en               | c. | ei-tje      | d. | eier-tje-s                | e. * | eier-tje               |
|         | egg   |    | egg <sub>ALL-PL</sub> |    | egg-DIM     |    | egg <sub>ALL-DIM-PL</sub> |      | egg <sub>ALL-DIM</sub> |
|         | 'egg' |    | 'eggs'                |    | 'small egg' |    | 'small eggs'              |      |                        |
- 
- |         |        |    |                          |    |              |    |                            |      |                         |
|---------|--------|----|--------------------------|----|--------------|----|----------------------------|------|-------------------------|
| (71) a. | blad   | b. | blader-en <sup>162</sup> | c. | blad-je      | d. | blader-tje-s               | e. * | blader-tje              |
|         | leaf   |    | leaf <sub>ALL-PL</sub>   |    | leaf-DIM     |    | leaf <sub>ALL-DIM-PL</sub> |      | leaf <sub>ALL-DIM</sub> |
|         | 'leaf' |    | 'leafs'                  |    | 'small leaf' |    | 'small leafs'              |      |                         |

Admittedly, these paradigms may sound slightly archaic to some speakers. Yet, crucially, it is very clear that even for these speakers the forms in the *e*-examples are illicit. The use of the allomorph in the singular diminutive is banned. This supports the claim that the allomorph is indeed determined by syntactic plurality for these examples.

Hermans and Postma (2009) argue that stem alternating words such as *kind* and vowel length alternating nouns belong to a specific noun class and that the additional morae in *kinder* and stems with long vowel express these noun classes in the presence of a syntactic plural head. Now, note, crucially, that vowel length alternating nouns never show the SSSL pattern, unlike the *kind-kind* alternation. It is therefore unlikely that they should be analysed on a par. More generally, it is unlikely that the distribution of allomorphs is regulated by the syntactic plural head if the non-diminutive and diminutive plural forms do not show the same allomorph. I therefore set aside a syntactic account for the allomorphic distribution of vowel length alternating noun in Dutch, even though I concur with Hermans & Postma (2009) that the analysis is on the right track for the *kind-kind* types. In short, syntactic plurality may determine allomorphy in Dutch. However, vowel length alternating nouns do not show this pattern.

## 8.6 Vowel alternating allomorphs in a split approach

In Section 2 I have argued that, from a theoretical point of view, a split approach according to which allomorphy selection is distributed over two modules, viz. Vocabulary Insertion and Phonology, is to be preferred. In section 3 to 5 I have discussed the facts for vowel alternating allomorphs in Dutch. I concluded that the data show a split pattern as well. I argued at length that considerable descriptive adequacy is to be lost if we do not recognise that Phonology is responsible for most of

---

<sup>161</sup> ALL is an abbreviation for allomorph.

<sup>162</sup> The vowel in *blader* is actually lengthened. Note that this allomorph contains the expected phonological pattern to select a lengthened vowel.

the data. Phonology alone allows us to understand the allomorph selection in non-diminutive nouns and verbs and it accounted for the absence of the unattested patterns. In the empirical domains of diminutive forms and morphologically derived forms, Phonology still allows us to capture many of the attested data. Yet, in these domains we have also encountered certain lexicalised exceptions. I argue to accept the data for what they are: the phonological pattern is real and the exceptions exist. The data echo opposing views in the literature: some emphasise on the phonological patterns (e.g. Wolf 2008), others warn that many examples are simply not phonologically optimal (e.g. Paster 2006). I argue, in contrast, that the opposing nature of these two types is not a theoretical embarrassment, but rather indicative of the way the selection of allomorphs is organised. More specifically, both Vocabulary Insertion and Phonology can determine the selection of allomorphs. Vocabulary Insertion is responsible for stored information, Phonology is responsible for phonologically optimising patterns. In this section I bring together these theoretical insights and the data that have been discussed in sections 3 and 4 and I illustrate how the mechanisms that I proposed in Section 2.9 allow us to derive the data.

Consider first a stem that shows nothing but phonologically expected behaviour, such as *hertog* ‘duke’, as shown in Table 3:

**Table 3:** the noun *hertog* ‘duke’

word	surface form	translation	type	phonologically expected?
hertog	[hərtɔx]	duke	singular noun	✓
hertogen	[hərtoyən]	dukes	plural noun	✓
hertogje	[hərtɔxjə]	small/cute duke	singular diminutive form	✓
hertogelijk	[hərtoyələk]	ducal	morphological derivation	✓

If we assumed a rich subcategorisation frame, as in Paster (2006), we basically would have to stipulate all information given in Table 3 under the relevant root at Vocabulary Insertion. The insight that this noun shows nothing but phonologically optimising behaviour is lost. However, under the split approach presented in Section 2.9, we could simply stipulate the following vocabulary item at Vocabulary Insertion:

$$(72) \quad \sqrt{4377} \leftrightarrow \{/hərtoy/, /hərtoy/\}$$

The vocabulary item states that the root has two exponents which are members of a disjunctive set. Given that Vocabulary Insertion does not state which one to use in a given context, it does not establish a preference. Note, crucially, that none of these forms is a default form: they are equal members of the set of options. Vocabulary Insertion thus absolutely fails to establish a preference. Both forms therefore serve as input to Phonology, which will determine which form is the most appropriate form. For all the reasons described in detail in the previous sections, Phonology knows exactly what to do: it will select the allomorph with the short vowel in closed syllables and the allomorphs with the long vowel in open syllables. Its selection results in predictable, optimised syllables.

Now consider a root which shows phonologically expected behaviour, except in one lexical context, as shown in Table 4:

**Table 4:** the noun god ‘god’

word	surface form	translation	type	allomorph phonologically expected?
god	[yɔt]	god	singular noun	✓
goden	[yodə(n)]	gods	plural noun	✓
godje	[yɔt̪ə]	small god	singular diminutive form	✓
afgoderij	[afyodərɛi]	idolatry	morphological derivation	✓
verafoden	[vərafyodə(n)]	idolise	verb	✓
goddelijk	[yɔdələk]	divine	morphological derivation	*

We want to make Phonology responsible for the selection of the allomorphs, unless the root is followed by the affix *-elijk*. This can be done by modelling the root as follows at Vocabulary Insertion:

$$(73) \quad \sqrt{692} \leftrightarrow /yɔd/ \ / \_\_ / -ələk / \\ \{/yɔd/, /yod/\}$$

The vocabulary item states that priority is given to the lexicalised fact that the vowel is short when the root precedes the affix *-elijk*. In all other contexts, the disjunctive set containing both allomorphs is sent as input to Phonology. At this point, we are making use of the advantage that stipulations at Vocabulary Insertion can enforce a phonologically less optimal form, as in Paster (2006). However, we are refraining from listing those forms that can be regulated by Phonology.

At risk of stating the obvious, in this model it is clear that Vocabulary Insertion precedes Phonology rather than the other way around. If we were to assume that Phonology precedes Vocabulary Insertion, a phonologically expected form would always win, as Phonology will rule out less optimal candidates. As I have pointed out in Section 2.8, the order presented here is also the only one that makes sense considering the design of the Y-model. The facts and the theoretical expectation thus go hand in hand. Furthermore, we do not have to stipulate a Control Mechanism that follows Phonology, as in Bye (2007), as all lexicalised selection is done by Vocabulary Insertion, the presence of which is postulated anyway in Distributed Morphology’s Y-model.

## 8.7 Conclusion

In this article I have argued that the selection of allomorphs is done by two separate mechanisms: Vocabulary Insertion selects learned occurrences of allomorphs, Phonology regulates optimising allomorphic patterns. As is generally assumed in Distributed Morphology’s Y-model, Vocabulary

Insertion precedes Phonology and the lexicalised selection is therefore given priority over phonological selection. I have argued that this proposal does justice to the observations in the literature that both types of allomorph selection seems to exist: allomorphy may be clearly phonologically optimising and it clearly may not be so in other cases.

I proposed a minor modification at Vocabulary Insertion to execute the idea technically. I proposed that Vocabulary Insertion may be marked for a set of unordered exponents. All exponents are equal candidates for insertion.

I have presented a case study from Dutch, i.e. vowel length alternating nouns, which shows that, indeed, these facts co-exist. Non-diminutive nouns and verbs with such allomorphs show phonologically expected behaviour. Furthermore, Phonology allowed us to understand the systematic absence of certain allomorphic patterns. In the domain of diminutive forms and morphological derivations the clear pattern broke down and we encountered exceptions, the allomorphic distribution of which is clearly non-optimising from a phonological point of view. I concluded that the phonological patterns are real and the exceptions are real. I have argued that both types result from different mechanisms: Vocabulary Insertion is responsible for the exceptions, Phonology is responsible for the phonological patterns.

# 9 Root allomorphy depends on head movement

## Support from Breton pluralisation

When marked for plural number, Breton nouns can surface as so-called simple or double plurals (Trépos 1957):

- |               |                  |                     |
|---------------|------------------|---------------------|
| (1)    merc'h | (2)    merc'h-ed | (3)    merc'h-ed-ou |
| ‘girl’        | ‘girls’          | ‘girls’             |

At first sight, these data look like instances of double exponence of plural marking (see Harris 2017 for a detailed discussion of the literature since Anderson 1986). Yet, I believe it is more likely that these examples are instances of root allomorphy for the following three reasons. Firstly, the stem allomorph is not inherently tied to plurality. It is also attested in derivations, as illustrated in (4). The root may get a non-plural interpretation in such derivations, as shown in the singulative in (5).

- |  |  |
|--|--|
| (4)    merc'het-a                        | (5)    stered-enn                      |
| girl <sub>ALLOMORPH</sub> -VERBAL.SUFFIX | star <sub>ALLOMORPH</sub> -SINGULATIVE |
| 'to womanize'                            | 'individual star'                      |

Secondly, the simple plural always *precedes* derivational suffixes (Stump 1989: 272). This is exactly what we expect; a stem allomorph cannot be broken apart by derivation. Thirdly, some ‘simple plurals’ are marked by means of a vowel modification in the stem, i.e. an ablaut. Whereas, admittedly, an allomorph such as *merc'hed* can still be analysed as a concatenation of the default root *merc'h* and a suffix *-ed*, such an analysis is not possible for an ablaut, which clearly shows that the stem is subject to allomorphy:

- |            |                          |                             |
|------------|--------------------------|-----------------------------|
| (6) louarn | (7) lern                 | (8) lern-ed                 |
| fox        | fox <sub>ALLOMORPH</sub> | fox <sub>ALLOMORPH-PL</sub> |
| 'fox'      | 'foxes'                  | 'foxes'                     |

In sum, the so-called simple plural is probably a stem allomorph that is inserted in the context of plurality.

Now observe the following facts. Cardinals and plural marking are in complementary distribution in Breton plural marking. Crucially, in the presence of a cardinal, the root allomorph cannot be selected and the default root surfaces obligatorily:



Discussing Turkish and Hungarian data, Borer (2005:116-117) provides the following analysis for the complementary distribution of cardinals and plural marking: in the absence of a cardinal, the Dividing (Div) feature is realised by the plural affix. When present, however, the cardinal realizes

both the feature [Div] (i.e. plural number in this case) and the [#]-feature, i.e. the feature regularly expressed by quantifiers which represents a counting function semantically. Consequently, in the presence of a cardinal, the noun (i.e. the root incorporated into little n or simply little n, depending on your analysis) does not head-incorporate into the plural morpheme, either syntactically (via head movement) or post-syntactically (via morphological merger). The data show that in the absence of such head movement, root allomorphy is blocked. These data confirm Bobaljik's (2012:68) hypothesis that root allomorphy depends on head movement: a head that conditions root allomorphy must be in the same complex head (i.e. the same morphological word) as the root.

# 10 The Extravagant Dutch Suffix *-ke* and its Meandering through the Interfaces

## Abstract

The substandard Dutch diminutive suffix *-ke* occurs in words of four different categories: nouns, predicatively used adjectives, adverbs and interjections. In nouns, adjectives and adverbs the suffix may trigger an epenthetic consonant, which is not available for the interjections. This seems to go hand in hand with a syntactic distinction: for the interjections the suffix is not to be analysed as a diminutive, but as an honorific. I argue that in all contexts, the suffix realises scalar features, but of different natures. The phonological contrasts follow from the extravagant meandering through the syntax-phonology interface.

## 10.1 Introduction

This contribution discusses the Standard Dutch diminutive suffix *-tje* and the substandard Dutch diminutive suffix *-ke*, both of which occur in words of four different categories: nouns, predicatively used adjectives, adverbs and interjections:

- |   |  |                |
|---|--|----------------|
| (1) een appel-tje<br>an apple-DIM(INITIVE)<br>'a small apple'                       | een appel-ke<br>an apple-DIM<br>'a small apple'                    | (noun)         |
| (2) Ze is ziek-je-s.<br>she is ill-DIM-S<br>'She is (somewhat) ill.'                | Ze is ziek-s-ke-s.<br>she is ill-DIM-S<br>'She is (somewhat) ill.' | (adjective)    |
| (3) Ze zong zacht-je-s. <sup>163</sup><br>she sang soft-DIM-S<br>'She sang softly.' | Ze zong zacht-eke-s.<br>she sang soft-DIM-S<br>'She sang softly.'  | (adverb)       |
| (4) Hallo-tje-s!<br>hello-DIM-S<br>'Hello!' (informal)                              | Hallo-ke-s!<br>hello-DIM-S<br>'Hello!' (informal)                  | (interjection) |

Furthermore, in certain phonological contexts, the substandard diminutive suffix *-ke* co-occurs with an epenthetic consonant, glossed as EC, in nouns (5) and adjectives/adverbs (6). In interjections, however, the epenthetic consonant is not available, and the examples are overall far less acceptable:

---

<sup>163</sup> I am unsure how to translate *zachtjes* adequately. To the best of my knowledge, no equivalent form exists in English. A native speaker will probably understand that she sang tenderly or shyly. Content-wise, it seems to me, the sentence is synonymous with 'Ze zong zacht'.

- (1) dag-\*(s)-ke (noun)  
 day-E(PENTHETIC)C(ONSONANT)-DIM  
 'short/nice day'
- (2) bleek-\*(s)-ke-s (adverb)  
 pale-EC-DIM-s  
 'slightly pale'
- (3) ?? dag-\*(s)-ke-s (interjection)  
 day-EC-DIM-s  
 Intended: 'Hi.'

The gloss 'diminutive' for the suffix *-ke* is in fact a misnomer for the example in (7). I will later argue that within interjections (such as 'hi') the suffix *-ke* does not function as a diminutive, but as an honorific. Honorifics are morphemes that express the semantics of (anti)politeness and they modify the degree of formality expressed by the interjection (see sections 6 and 7 for details).

Dutch diminutive suffixes occur in three different categories. As such, it is extravagant in that it is an instantiation of categorial unfaithfulness. The availability of an epenthetic consonant that is, *nota bene*, dependent on *phonological* contexts is limited to two of the categories. We will furthermore see that both the diminutives and honorifics are sensitive to the phonological Obligatory Contour Principle (Leben 1973), albeit for different features. These facts point to extravagant behavior at the interfaces, to which I come back below (insert ref to section?).

The use of the suffix in the interjections is quite innovative. It mostly belongs to the sociolect of the younger generations. It may be considered as a form of expressive language use. As such, it is certainly an instance of an affix gone *extravagant*, particularly since the mere fact of interjections containing honorific affixes is quite unusual.

The data certainly contribute to the study of extravagance through interface meandering (we will insert a cross-ref to the intro later). We will see that the occurrence of the suffix in these three contexts (nouns, adjectives/adverbs, interjections) is justified by the fact that the three contexts crucially share the semantic property of marking a low value on a scale, which I assume to be represented in the morphosyntactic structure. The lexicon then allows the suffix to realise the scale head and the features contained on this head by a single suffix for all three contexts. When it comes to the epenthetic consonant, I will show that its occurrence is clearly conditioned phonologically. However, the phonological domains are also syntactically determined. The affix and the stem need to be in the same cyclic domain<sup>164</sup> in the syntactic structure to be spelled out together and thus to form a determining phonological domain. It will become clear that we can capture a great deal of the empirical observations from an intricate interplay between various modules: semantics, syntax, lexical insertion and phonology. However, we will have to conclude that we cannot fully account for the differences in the Obligatory Contour Principle. As such, the aim of this

---

<sup>164</sup> A cyclic domain is also called a phase. It refers to the hypothesis that a (morpho)syntactic structure is not first built in its entirety and then spelled out to the phonological form and the logical form, but rather built per cyclic domain, i.e. per specific units of structure (Chomsky 2001, Marantz 2007). Structure-building is thus cyclic: a bit of structure is built and spelled out, then a higher piece of the structure is built and spelled out and so on. Spell-out is triggered by phase heads. However, which heads exactly are supposed to be phase heads is subject to debate.

article is to define a puzzle for the syntax-phonology interface from Dutch and to hint at some directions to solve the puzzle.

The substandard nouns with the diminutive suffix *-ke* are well-known in the literature and have been described precisely (see e.g. De Schutter et al. 2005: 35-50). The interjections with the diminutive suffix, on the other hand are less described forms which are restricted both by geographical factors, register (informal, spoken Dutch) and social factors (younger generations). The interjections presented in this paper were collected by me in 2014 in Belgian Dutch. They were judged by native speaker 2<sup>nd</sup> year students ( $n = 7$ ) who followed a course in Dutch phonology and morphology that I taught in 2014 at the KU Leuven campus Brussels.

My contribution is structured as follows. Section 0, , I focus on the epenthetic consonant as it occurs in the nominal domain. Section 4 discusses how this epenthetic consonant fits within the typology of epenthetic consonants. Section 5 presents a comparison with Yiddish, from which it will become clear that the epenthetic consonant is restricted to a specific syntactic domain. Section 6 argues that the epenthetic consonant is indeed absent in a different syntactic domain, viz. in honorifics. I will then complete the picture by pointing out that it is present, however, in adjectives and adverbs. These observations form the empirical basis for my analysis. It will become clear that all syntactic domains share some specific features. However, the honorifics differ syntactically from the others in their cyclic domains, resulting in the non-availability of the epenthetic consonant. In the conclusion I point out that some puzzling aspects of these data can be captured, but I also point at a remaining unresolved issue.

## 10.2 Distributed Morphology

The present study is grounded within the framework of Distributed Morphology (Halle and Marantz 1993, 1994; Harley and Noyer 1999) and it argues for certain views on the interfaces that are indeed dominant within this framework. There are four properties of Distributed Morphology that are crucial to understand the present contribution: morphology is syntax, Late Insertion, roots and cyclic structure-building. These properties will be discussed in turn.

The idea that morphology is syntax involves the hypothesis that there is only a single structure building module in the human faculty of language, called syntax, which builds both words and sentences. Morphology and syntax are thus not assumed to be separate modules. Words are simply the lowest parts of structures in syntactic tree.

Late Insertion builds on the assumption that syntax operates with abstract features rather than lexical items. For example, rather than merging the definite article *the* in the tree, the syntactic features [definite] and [determiner] will be merged on a terminal node in the tree. Lexical insertion takes place post-syntactically, hence ‘Late Insertion’. Vocabulary items are marked with features, as in the following examples, which represent the articles in Dutch:

- (4)
- |            |                   |   |
|------------|-------------------|---|
| <i>een</i> | $\leftrightarrow$ | {determiner, indefinite, countable: singular} |
| <i>ø</i>   | $\leftrightarrow$ | {determiner, indefinite}                      |
| <i>de</i>  | $\leftrightarrow$ | {determiner, definite}                        |
| <i>het</i> | $\leftrightarrow$ | {determiner, definite, neuter, singular}      |

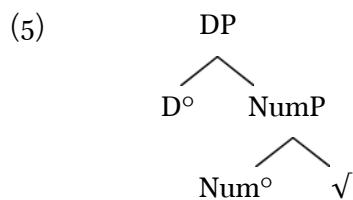
The vocabulary item of which the features are the closest match to the features of the syntactic node will then realise the syntactic node. For example, if the syntactic node is marked with the

features {determiner, definite, neuter, singular} the vocabulary item *het* is the closest match (in fact, an exact match in this case) and it will realise the syntactic node. The closest match is determined via the Subset Principle. As a result, the features of the lexical item can be an exact match for the features of the terminal node (as in the example just mentioned) or it can be a subset thereof. For example, a syntactic node which is marked for the features {determiner, definite, common} will be realised by the vocabulary item *de*, even though *de* is not marked for common gender.<sup>165</sup> When several lexical items compete for insertion, the lexical item which shares the most features with the syntactic node will be inserted. A lexical item that is marked with a feature that is not present in the syntactic node will not be inserted (i.e. supersets are not eligible for insertion).

The lowest node in the syntactic tree is the root: the root node. This node will be realised by a lexeme, which is also called a root.<sup>166</sup> Root nodes are assumed to be devoid of any syntactic features. It follows that they lack a categorial feature as well. Indeed, roots are taken to be categoriless: the category of the structure is derived from the functional structure that merges on top of the root. For example, a root with nominal superstructure will be interpreted as nominal, one with verbal structure will be interpreted as verbal, etc.

Cyclic structure-building refers to the idea that the syntactic structure is built in cycles or phases (Chomsky 2001, Marantz 2007). It is not assumed that an entire clause is built before it is sent to the phonological and semantic interfaces. The theoretical hypothesis is that subparts of the clausal structure form units of Spell-out. While the exact nature of these units is under debate, it is commonly assumed that categorial heads, which are realised by derivational affixes, are phase heads and as such they trigger Spell-out at the interfaces.

Following Borer (2013a) and deviating from Distributed Morphology (DM), I assume that the system lacks productive null categorisers which would categorise a root, i.e. by nominalising or verbalising it. Rather, I assume that the category of the root is rather derived from the functional superstructure, such as a Voice head, a transitive head or an agentive head for verbs, and plurality or a determiner for nouns (for an extensive demonstration of a DM approach without null categorisers, see De Belder 2011a). This is illustrated the nominal structure in (9). The symbol √ refers to the root which is nominalised by merging with inherently nominal superstructure, namely NumP, which encodes number, and DP, which introduces the determiner.



The present contribution aims at illustrating how extravagant morphology is captured in Distributed Morphology. We will see that extravagant behavior follows from the Subset Principle and the fact that vocabulary items can be underspecified for a syntactic context. As such, a vocabulary item can be somewhat flexible in the meaning it expresses and it may show some degree of categorial unfaithfulness. Extravagant behavior through interface meandering follows from cyclic

<sup>165</sup> The article *de* is not restricted to contexts of common gender. It is also used for neuter nouns in plural contexts. The fact that *de* is therefore underspecified for gender is expressed by leaving out a gender feature on the marking of the vocabulary item.

<sup>166</sup> In other frameworks, roots are known as lexemes or open class vocabulary.

Spell-out: pieces of different syntactic size may be sent to the phonological and semantic interfaces, resulting in different patterns.

### 10.3 The epenthetic consonant

I will start with the Dutch suffix *-ke*, focusing on the epenthetic consonant [s] that occurs in the Dutch nominal diminutive forms, as shown in (9). I thus postpone a discussion of the adverbs, adjectives and the interjections to later sections.

- (6) dag-\*<sup>(s)</sup>-ke [daxskə]  
 day-EC-DIM  
 'short/nice day'

The epenthetic consonant only occurs when a diminutive form with the suffix *-ke* is formed. In contemporary Dutch, this diminutive form belongs to substandard language use. It is commonly attested in Dutch dialects, most notably in the Southern and Eastern dialects.<sup>167</sup> The data below are Southern Dutch (i.e. Belgian Dutch) (De Schutter et al. 2005). The suffix *-ke* is etymologically related to the German diminutive suffix *-chen* and the older English suffix *-kin*, which can still be recognised in words such as *pumpkin* and *napkin*.

The [s] only appears after velar consonants. Yet, in this context, it is obligatory:

- (7) a. tak-\*<sup>(s)</sup>-ke [k]  
 branch-EC-DIM  
 'small branch'  
 b. zeg-\*<sup>(s)</sup>-ke [x]  
 say-EC-DIM  
 'short opiniated verbal contribution'  
 c. ring-\*<sup>(s)</sup>-ke [ŋ]  
 ring-EC-DIM  
 'small/nice ring'

It does not occur in other phonological contexts:

- (8) a. Marij-(\*s)-ke [ɛɪ]  
 Mary-EC-DIM  
 b. maan-(\*s)-ke [n]  
 moon-EC-DIM  
 c. vrouw-(\*s)-ke [w]  
 woman-EC-DIM  
 d. aap-(\*s)-ke [p]  
 monkey-EC-DIM  
 e. duif-(\*s)-ke [f]  
 pigeon-EC-DIM

---

<sup>167</sup> The Standard Dutch affix *-je* is most probably derived from *-ke* (De Schutter et al. 2005: 34).

As the epenthetic consonant only occurs after velar consonants, it is certainly triggered by its phonological context. This phenomenon can be ascribed to the Obligatory Contour Principle, which states that certain consecutive identical features are prohibited in phonological representations (Leben 1973, Goldsmith 1976, Gussenhoven and Jacobs 1998: 158-161, see Pierce & Boas 2010: 214 for discussion). Given that the velar consonant /k/ of the diminutive suffix is banned from following a root ending in a velar consonant, the problematic identical feature in this case is the place feature [velar]. The epenthetic consonant [s] thus improves the forms by interrupting a sequence of velar consonants.<sup>168</sup>

Note that velar consonants are not generally banned from following one another in Dutch. The epenthetic consonant does not occur in different syntactic structures with a comparable segmental context:<sup>169</sup>

- (9) Dag Kerensa [daxkərənsa]  
 hi Kerensa  
 'Hi, Kerensa!'

The epenthetic consonant therefore does not only depend on a specific segmental context alone. Phonology thus does not suffice to capture its distribution. To understand the behavior of this epenthetic consonant, the next section describes typological insights into various types of epenthetic consonants.

#### 10.4 The [s] is a syntactic epenthetic consonant typologically

Zygis (2010) proposes a typology of epenthetic consonants. She distinguishes between the following three types: grammatical (i.e. morphosyntactic) epenthetic consonants, phonetic epenthetic consonants and prosodic ones.

The German example in (14) shows a grammatical epenthetic consonant (example taken from Zygis 2010: 112, her example (2)):

- (10) wesen-t-lich [German]  
 stem-ec-suffix  
 'considerably'

For grammatical epenthetic consonants, the epenthetic consonant is triggered by morphosyntax. In (14), the suffix *-lich* triggers the presence of the epenthetic [t]. Zygis (2010) notes that such consonants are often coronal sounds, as is illustrated in the example. Such epenthetic consonants are language-specific. Indeed, the suffix *-lich* is a German suffix and the fact that it triggers the epenthetic consonant is indeed a German-specific phenomenon. Such epenthetic consonants are

---

<sup>168</sup> Some dialects allow for the insertion of a schwa in between the two velar Cs. The epenthetic consonant still appears even though the velar Cs are not adjacent: *dageske* [daxəskə], *ringeske* [rɪŋəskə].

<sup>169</sup> Note that such contexts *de facto* boil down to phonological word boundaries, given the inventory of affixes in Dutch and the anti-gemination rule.

obligatory and, typologically, they may or may not depend on the segmental context, as noted by Zygis (2010: 112).<sup>170</sup>

The following two examples illustrate the use of phonetic epenthetic consonants. ((15) is taken from Zygis (2010: 112, her example (2)):

- (11) sytu(w)acja [Polish]  
'situation'

- (12) Pa(j)ola [Sint-Niklaas Dutch]  
'Paola'

Such epenthetic consonants are fully determined by phonetic, articulatory factors. This is concluded from the fact that similar data are found in unrelated languages. The phonetic epenthetic consonant is often a stop or a glide and it is sensitive to speech rate. It is fully determined by the segmental context.

A prosodic epenthetic consonant can be seen in (17):

- (13) (?)aapa inni [Selayarese]  
'What is this?'

Such epenthetic consonants are conditioned by prosodic boundaries. They are, however, mostly independent of the segmental context. They are often stops or fricatives, and one may find similar data in unrelated languages. Within a language, they may be subject to variation.

It may be clear that the epenthetic consonant [s] that occurs in the Dutch diminutive observes all the criteria of a grammatical epenthetic consonant. We have seen above that it is specifically triggered by a specific affix, i.e. the diminutive suffix *-ke* and as such it is a phenomenon that is specific to Dutch.<sup>171</sup> It is a coronal sound, and in the relevant context, it is indeed obligatory and not subject to the speech rate or variation. We have seen that it is determined by its segmental context as well, which is possible for such epenthetic consonants. In conclusion, the epenthetic consonant under discussion should be classified as a syntactic epenthetic consonant.

One may wonder what exactly is meant by stating that the epenthetic consonant is triggered solely by the diminutive exponent *-ke*. Does it mean that it is the vocabulary item, i.e. the specific exponent, that determines the occurrence of the epenthetic consonant? Or does it mean that the syntactic features also determine its occurrence? The question has immediate descriptive relevance. Does the epenthetic consonant occur whenever we see the suffix *-ke* or does it occur whenever we see *-ke* in a specific syntactic structure?

This question is particularly relevant in the framework of Distributed Morphology, as it is a Late Insertion model. The specific vocabulary item *-ke* will only be matched to the structure post-syntactically at a module called vocabulary insertion. The abstract syntactic tree and the tree as realised by vocabulary items are thus clearly distinguished in this model.

---

<sup>170</sup> An anonymous reviewer notes that for the German example in (14) the epenthetic consonant seems to be triggered by the final segment, /n/, in *wesen*.

<sup>171</sup> In the next section, it will become clear that we observe similar data in Yiddish, but it still holds that the epenthetic consonant is specific here to a group of languages, i.e. West-Germanic languages.

In the following, I will argue that it is the underlying syntactic structure that triggers the epenthetic consonant, not the specific affix. To do so, I will address data from Yiddish in the next section that show that an epenthetic consonant occurs in the same syntactic context as in Dutch, but with a very different exponent. In later sections I will present Dutch data which show the same exponent, but a different syntax. It will become clear that the epenthetic consonant does not occur in these data.

### 10.5 Same syntax, different lexical items

In this section I present data that show that an epenthetic consonant occurs in Yiddish in the same syntactic structure as it does in Dutch, even though it has a different phonological context. More specifically, in Yiddish one may observe an [x] after a non-syllabic [l] and a [d] after an [n] (Pierce & Boas 2010).

- (14) *mayl-x-l*

mouth-EC-DIM

‘small mouth’

- (15) *beyn-d-l*

bone-EC-DIM

‘small leg/bone’

- (16) *kastn-d-l*

closet-EC-DIM

‘small closet’

As in Dutch, the epenthetic consonants only appear in the diminutive in Yiddish (Pierce & Boas 2010). Pierce & Boas (2010) note that a sequence of /nl/ is not illicit in Yiddish, as can be seen in (21). The pair in (22) and (23) shows a minimal contrast morphologically and phonologically. Both are instances of suffixation and both show the same relevant phonological context. Yet, only the form with the diminutive affix triggers the epenthetic consonant.

- (17) *finland*

‘Finland’

- (18) *payn-lex*

pain-SUFFIX

‘painfully unpleasant’

- (19) *beyn-d-l-ex*

bone-EC-DIM-PL(URAL)

‘small legs/bones’

Pierce & Boas (2010: 214) analyse these data as being subject to the Obligatory Contour Principle, a proposal that I assumed for Dutch as well, as discussed above.

It should be clear that the phenomenon does not involve the same lexical items in Dutch as it does in Yiddish. In Dutch, the epenthetic consonant co-occurs with the lexical item *-ke*,<sup>172</sup> in Yiddish it involves the lexical item *-l*. One might stipulate that, coincidentally, both the Yiddish *-l* suffix and the Dutch *-ke* suffix share the idiosyncratic property of requiring an epenthetic consonant. Alternatively, and more interestingly, one might hypothesise that it is a syntactic feature in the diminutive structures of these related languages that triggers the epenthetic [s]. We will proceed with the following hypothesis: The epenthetic consonant in West-Germanic diminutives is triggered by an underlying syntactic feature, not by a specific property of a lexical item. Support for this hypothesis comes from the following section, which shifts the attention to the honorifics.

## 10.6 Same lexical item, different syntax

### 10.6.1 The suffix *-ke* may occur as an honorific

As we have noted above, substandard spoken Belgian Dutch has a diminutive-like suffix *-ke* that attaches to interjections:

- |  |   |   |
|--|---|---|
| (20) a. merci-ke-s<br>thank.you <sup>173</sup> -KE-s <sup>174</sup><br>'thank.you' | b. hallo-ke-s<br>hello-KE-s<br>'hi'                 | c. ciao-ke-s<br>ciao-KE-s<br>'bye'  |
| d. amai-ke-s<br>INTERJECTION-KE-s<br>'oh my god'                                   | e. bravo-ke-s<br>bravo-KE-s<br>'bravo'              | f. hoera-ke-s<br>hurray-KE-s<br>'hurray'                                  |
| g. salu-ke-s<br>bye-KE-s<br>'bye'  | h. sorry-ke-s<br>sorry-KE-s<br>'sorry'              | i. godverdomme-ke-s<br>goddamn-KE-s<br>'goddamn'                          |
| j. elaba-ke-s<br>hey.there-KE-s<br>'hey there'                                     | k. oei-ke-s<br>oy-KE-s<br>'oy vey'                  | l. hola-ke-s<br>INTERJECTION <sup>175</sup> -KE-s<br>(expresses surprise) |
| m. shalom-ke-s<br>shalom-KE-s<br>'bye'   | n. aw-ke-s<br>INTERJECTION-KE-s<br>(expresses pain) | o. miauw-ke-s<br>meow-KE-s<br>(expresses sexual interest)                 |

<sup>172</sup> All morphemes count as lexical items. Affixes are thus instances of lexical items.

<sup>173</sup> *thank.you* may or may not be an ideal gloss: a reviewer wonders whether *thank you* is indeed an interjection. This discussion would take us too far afield and I would like to avoid it here. The Dutch word *merci* (borrowed from French), in any case, is used as an interjection in Dutch.

<sup>174</sup> The reader may wonder what the function of the *-s* is, which occurs on interjections and adverbs in Dutch. It has been studied by Corver (2017), who argues that it invariably spells out a categorial head. I refer readers to his work for more details.

<sup>175</sup> When the interjection does not translate into English, I merely gloss them as 'interjection'.

The meaning of this morpheme is solely expressive: it expresses a high degree of informality with the addressee. The informality adds a positive tone. As a side-effect the *-ke* morpheme may soften down the tone or it may be humorous.

I propose that these examples should be analysed as honorifics. I define honorifics as morphemes that express the semantics of (anti-)politeness, which may refer to varying degrees of social distance, psychological distance and formality (see below where I discuss the work of McCready 2019). Several arguments support this proposal. First, consider the fact that, cross-linguistically, honorifics may be related to diminutives. For example, the Japanese suffix *-chan* is an honorific for children, pets, females and cute people. Its meaning is comparable to diminutives: it expresses smallness, endearment or cuteness.

- (21) Maki-chan (Japanese)  
proper.name-honorific

Secondly, honorifics may express informality cross-linguistically. Korean, for example, has various speech styles. The *banmal* style expresses informality and intimacy. Similarly, the Indo-European *tu* vs. *vous* distinction is partly based on a distinction in (in)formality and (non)familiarity.

Thirdly, Dutch interjections can get honorifics other than the *-ke* suffix. For example, Blankenberge Dutch may distinguish between greetings addressed to strangers and greetings addressed to people one is acquainted with:<sup>176</sup>

- (22) a. Hey!  
(greeting, neutral)  
b. Heeey! (dragged vowel)  
(greeting, obligatorily implies acquaintancedness)

Addressing a stranger in Blankenberge using the form with the dragged vowel is a social faux pas. Therefore, I propose that the lengthening is a lexical item realising an honorific (cf. Postma 2001 who proposes that a dragged vowel may realise negative polarity in Dutch). Note, for good measure, that the lengthened vowel differs from the *-ke* morpheme as it implies acquaintancedness rather than informality:

- (23) a. Hey!  
(greeting, neutral) b. Heykes!  
(greeting, highly informal)

Fourthly, note that honorifics may attach to interjections cross-linguistically. The honorific attached to the greeting in the following Kashmiri example expresses non-acquaintedness (Koul *sine dato*: no page numbers):

- (24) aadaab arIz jinaab, myeny kAAm kArytav haz  
greetings HON(ORIFIC), my work do-IMP(ERATIVE)POLITE.HON  
'Greetings, please do my work.' (Kashmiri)

---

<sup>176</sup> The phenomenon is not restricted to this dialect.

Fifthly, the morpheme *-ke* can attach to a personal pronoun. A specific Dutch sociolect allows *-ke* on 2<sup>nd</sup> person plural pronouns:

- (25) a. jullie  
          you.PL

b. u  
    you.FORMAL

c. jullie-ke-s  
    you.PL-INFORMAL.HON-S

For good measure, it should be noted that this use of the honorific on this pronoun is restricted to the most informal use of Belgian spoken Dutch and it is quite rare. However, it can be observed (see footnote 21).<sup>177</sup> Whereas honorifics on interjections are rather rare cross-linguistically, they are common on personal pronouns. Given these observations, I conclude that the suffix *-ke* realises an honorific on interjections.

#### 10.6.2 The honorific and the diminutive are realised by the same lexical item

One might suspect that Dutch has two homonymous lexical items *-ke*, a diminutive suffix and an honorific suffix. Alternatively, one may suspect there is a single suffix that is underspecified enough to be inserted in both syntactic contexts. This second hypothesis is the most likely one for the following reason. Dutch has two diminutive suffixes: *-ke* and *-tje*, which is used in Standard Dutch. Both the substandard suffix *-ke* and the Standard Dutch suffix *-tje* have allomorphs:

- (26) -ke allomorphs: -eke  
(27) -tje allomorphs: -etje, -pje, -je

Both diminutive suffixes may occur in honorific contexts. Importantly, the allomorphy of the exponents in the DP and the exponent on the interjections co-vary: if the phonological context requires the use of a specific allomorph in the diminutive, the same phonological context will select the same allomorph in the honorific:



<sup>177</sup> Here is a googled example which combines *heykes* and *julliekes* in the same sentence: *piep héykes* (hey-INFOR-MAL.HON-S) (*sic*) *rob en amber dit filmpje was totaal wat ik van julliekes* (you.PL-HON-S) *verwachte* (*sic*) *doe zo verder zou ik zeggen !!!* (“Peep, hey, Rob and Amber, this film was totally what I expected from you guys. Do continue like this, I would say.”) ([www.lipsticksex.net](http://www.lipsticksex.net), Naughtycouple, [17/8/2004]).

- (29) a. do-tje  
do-DIM  
'short do'
- b. hallo-tje-s (informal Standard Dutch)  
hi-KE-s  
'Hi!' (informal)
- (30) a. boom-ke  
tree-DIM  
'small tree'
- b. shalom-ke-s (substandard Belgian Dutch)  
shalom-KE-s  
'Hi!' (informal)
- (31) a. boom-eke  
tree-DIM  
'small tree'
- b. shalom-eke-s (Antwerpen Dutch)  
shalom-KE-s  
'Hi!' (informal)
- (32) a. boom-pje  
tree-DIM  
'small tree'
- b. shalom-pje-s  
shalom-KE-s  
'shalom'
- c. salaam-pje-s (informal Standard Dutch)  
salaam-KE-s  
'salaam'

This observed co-variation is immediately captured if one accepts that the exponents in the DP and the ones on the interjections are invariably one and the same lexical item.

### 10.6.3 Restrictions on productivity

The honorific morpheme on interjections is highly productive in the informal spoken language of the younger generations of Dutch speakers in Belgium. Yet, this productivity is not unrestricted, as shown by the illicitness of the following examples:

- (33) a. \* njet-ke-s  
not-HON-s
- b. \* danku-ke-s  
thank.you- HON-s
- c. \* hosannah-ke-s  
hosannah-HON-s

There are two factors that restrict this productivity. Let us start with the trivial observation that the honorific is highly sensitive to register, as can be expected. It only attaches to interjections that are part of informal, spoken, Belgian Dutch vocabulary. Unsurprisingly, it does not attach to interjections of a higher register:

- (34) a. \* danku-ke-s  
thank.you-HON-s
- b. \* aangenaam-ke-s  
nice.to.meet.you-HON-s
- c. \* hosannah-ke-s  
hosannah-HON-s

The second factor concerns an observation that is more relevant to the present discussion and which involves a phonological restriction. In honorifics, the suffix *-ke* prefers roots whose coda has a high sonority. Most often it selects vowels (including the schwa) as in most examples below (e.g. *merci* ends in the vowel *i*), but it may attach to semi-vowels /j/ and /w/ as well (as in *oei* [uj] and *aw* [aw] below). Examples with sonorant consonants (such as the /m/ in *sjalom*) are more rare, but not excluded:

- |   |   |   |
|---|---|---|
| (35) a. merci-ke-s<br>thank.you-HON-s<br>'thanks' | b. hallo-ke-s<br>hello-HON-s<br>'hi'    | c. ciao-ke-s<br>ciao-HON-s<br>'bye'               |
| d. amai-ke-s<br>interjection-HON-s<br>'oh my god' | e. bravo-ke-s<br>bravo-HON-s<br>'bravo' | f. hoera-ke-s<br>hurray-HON-s<br>'hurray'         |
| g. salu-ke-s<br>bye-HON-s<br>'bye'                | h. sorry-ke-s<br>sorry-HON-s<br>'sorry' | i. godverdomme-ke-s<br>goddamn-HON-s<br>'goddamn' |
| j. elaba-ke-s<br>hey.there-HON-s<br>'hey there'   | k. oei-ke-s<br>oy-HON-s<br>'oy vey'     | l. hola-ke-s<br>interjection-HON-s<br>'oh wait'   |

- |  |  |   |
|--|--|---|
| m. aw-ke-s<br>INTERJECTION-HON-S<br>(expresses pain) | n. miauw-ke-s<br>meow-HON-S<br>(expresses sexual interest) | o. shalom-ke-s<br>shalom-HON-S<br>'bye' |
|--|--|---|

The suffix *-ke-s* is rare and slightly marked, but not excluded after the non-velar fricatives /s/ and /f/:

- |   |   |
|---|---|
| (36) a. ? helaas-ke-s<br>alas-HON-S<br>'alas' | b. ? oef-ke-s<br>INTERJECTION-HON-S<br>(expresses relief) |
|---|---|

But the suffix does not attach to plosives<sup>178</sup> or velars productively:

- |                                       |                                  |                                |
|---------------------------------------|----------------------------------|--------------------------------|
| (37) a. *? lap-(e)ke-s<br>shoot-HON-S | b. *? not-(e)ke-s<br>nope-HON-S  | c. *? hup-(e)ke-s<br>hup-HON-S |
| d. *? koest-(e)ke-s<br>hush-HON-S     | e. *? shit-(e)ke-s<br>shit-HON-S |                                |
| (38) a. * ach-ke-s<br>oh.well-HON-S   | b. * dag-ke-s<br>bye-HON-S       | c. * kak-ke-s<br>shit-HON-S    |

It seems that we see, yet again, the Obligatory Contour Principle at work. However, rather than solely being sensitive to a place feature, i.e. velar, it also observes manner this time: plosives are excluded.

Inserting the epenthetic consonant [s], only slightly improves the situation, as illustrated in (43) for (4ob). Yet, they are still ungrammatical and less acceptable than one expects them to be on purely phonological grounds, as is clear from a comparison with *helaas* 'alas' (see 44), an interjection ending in an [s]:

- |   |                                    |
|---|------------------------------------|
| (43) a. *? ach-s-ke-s<br>oh.well-EC-HON-S | b. *? dag-s-ke-s<br>bye-EC-HON-S   |
| c. *? kak-s-ke-s<br>shit-EC-HON-S         | d. *? fuck-s-ke-s<br>fuck-EC-HON-S |

- (44) ? helaas-ke-s  
alas-HON-S  
'alas'

The following pairs are telling in this respect: they are fully acceptable as diminutives in a DP, but unacceptable as interjections:

---

<sup>178</sup> There is one attested (?) example of -(e)kes after a plosive (*ochgottekkes* 'oh-god-DIM-S' (expresses empathy)).

- (39) a. twee dag-s-ke-s  
          two day-EC-DIM-PL  
          'two days'

- b. \*? dag-s-ke-s  
          day-EC-HON-s  
          Intended: 'Hi!'

- (40) a. twee kak-s-ke-s  
          two shit-EC-DIM-PL  
          'two turds'

- b. \*? kak-s-ke-s  
          shit-EC-HON-s  
          Intended: 'Shit!'

- (41) a. twee fuck-s-ke-s  
          two fuck-EC-DIM-PL  
          'two fucks'

- b. \*? fuck-s-ke-s  
          fuck-EC-HON-s  
          Intended: 'Fuck!'

I therefore conclude that the epenthetic consonant [s] is not available for honorifics on interjections. If it is correct that the epenthetic consonant is triggered syntactically by the diminutive and not specifically by the lexical item *-ke*, it follows that a different syntactic context, viz. the honorific, does not trigger it.

Before I conclude this section, I would like to add a short remark. One may have noticed that the honorific affix favors borrowed interjections. I believe this is a side-effect of the two restrictions observed above. Firstly, there is the fact that loaned interjections belong to an informal register. To wit, *hosannah* is a loanword as well, yet it does not select this honorific. Simply postulating that loanwords select the honorific would thus be the false generalisation. Secondly, there is the fact that loaned interjections may more readily end in a long vowel, which is actually quite rare in heritage Dutch phonology. As such, these loaned forms form an ideal phonological context of high sonority for the affix.

To summarise, the exponent realising the honorific is the same lexical item as the exponent realising the diminutive. Yet, the two affixes differ syntactically: the one that appears on the interjections is an honorific, not a diminutive. The diminutive selects the epenthetic consonant in the relevant phonological context, whereas the honorific does not. It is thus not the lexical item, but the syntactic context which determines the occurrence of the epenthetic consonant in diminutives. Furthermore, the phonological requirements of the lexical item *-ke* differ depending on the syntactic context. For the diminutives, the Obligatory Contour Principle only observes place features, whereas for the honorifics it observes both place and manner features.

## 10.7 The path through the interfaces

The present section aims to formalise the proposal that two different syntactic heads (a diminutive and an honorific) are realised by means of the same affix and that the epenthetic consonant depends on the syntactic context. If a single lexical item *-ke* realises both diminutives and honorifics it should be possible to determine which shared feature exactly is realised by this affix. Furthermore, it should be possible to determine which properties of the syntactic structure exactly regulate the occurrence of the epenthetic consonant.

### 10.7.1 The features realised by *-ke/-tje*

This far, I have described an opposition between diminutives in nominal groups and honorifics on interjections, and I have argued that the epenthetic consonant is restricted to the diminutives in the nominal groups. One might therefore suspect that the epenthetic consonant is somehow inherently nominal. However, this assumption would be false since the epenthetic consonant is not restricted to nominal contexts. It also occurs on adverbs and predicatively used adjectives which are (derived from) gradable adjectives:

- (42) a. ziek-s-ke-s<sup>179</sup>

ill-EC-DIM-s

'slightly ill'

- b. bleek-s-kes

pale-EC-DIM-s

'slightly pale'

Note that comparable forms are formed with the Standard Dutch diminutive form:

- (43) a. ziek-je-s

ill-DIM-s

'slightly ill'

- b. bleek-je-s

pale-DIM-s

'slightly pale'

Unlike, for example, English (*-ly*) or French (*-ment*), Dutch has no default overt affix that realises an adverbial head. Adjectives can simply be used as adverbs. However, the diminutive cannot occur on all forms. It is restricted to forms that show no overt agreement with the noun. As such, it will not occur on attributively used adjectives, which obligatorily agree with the noun, but it does occur on predicatively used adjectives and adverbs.<sup>180</sup>

---

<sup>179</sup> Readers may be familiar with the attributive-predicative opposition of adjectival inflection from German, which shows the same pattern as Dutch.

<sup>180</sup> The suffix *-s* occurs on interjections and adverbs in Dutch. As pointed out before, for details, the reader is referred to Corver (2017), who argues that *-s* invariably spells out a categorial head.

The diminutives, the honorifics and the adverbs should have a feature in common that is realised by *-ke*. There should also be a relevant distinction that sets apart the diminutives and the adverbs on the one hand, and the interjections on the other, such that the epenthetic consonant is restricted to the diminutives and adverbs.

In the following, I will show that the three data share one property: they all involve a scale.

De Belder (2011: chapter 3) argues in detail that the diminutive introduces a scale of size and implies a low value on a scale of size: the root merges with a measure function that implies that the root refers to a measurable unit. The diminutive then indicates that the value on this scale is a low value, i.e. a value of smallness. For example, the root *appel* ‘apple’ merges with the *Size<sup>o</sup>* head which guarantees that the apple is measurable. In other words, it turns the apple into a countable unit<sup>181</sup> (setting it apart from mass readings or kind readings). As a result, a diminutive form is excluded in mass readings (50) and kind readings (52). The examples are fully licit without the diminutive form, see (51) and (53).

- (44) \* Ze at appel-tje.  
she ate apple-DIM

- (45) Ze at appel.  
she ate apple  
'She ate some apple.'

- (46) \* Ze bestudeerde twee appel-tje-s: de Jonagold en de Granny Smith.  
she studied two apple-PL: the Jonagold and the Granny Smith

- (47) Ze bestudeerde twee appel-s: de Jonagold en de Granny Smith.  
she studied two apple-PL: the Jonagold and the Granny Smith  
'She studied two kinds of apples; the Jonagold and the Granny Smith.'

It further indicates that on the standard measure of apples, this particular unit scores rather low.

- (48) een appel-tje  
an apple-DIM  
'a small apple'

For the analysis of the gradable adverbs, we can borrow insights from the literature on the semantics of gradable adjectives. I believe this is justified as these adverbs are directly derived from such adjectives. It is well-known that gradable adjectives are traditionally analysed as involving a scale (Cresswell 1976, Kennedy 1999). For example, the adjectival group *very tall* implies that there is a scale of tallness and that the implied degree far exceeds the standard degree of tallness. Similarly, the *-ke* affix in the adverbs refers to a degree on a scale. For the examples below, the relevant scales are illness and paleness and again it is implied that the degree on this scale is rather low.

---

<sup>181</sup> Recall that roots are not marked for categorial features in Distributed Morphology. The noun *appel* ‘apple’ is thus not a count noun by itself. The root depends on the syntactic structure to acquire the meaning ‘countable noun’.

- (49) a. ziek-s-ke-s  
          ill-EC-DIM-ADVERBIAL.S  
          'slightly ill'

b. bleek-s-kes  
       pale-EC-DIM.ADVERBIAL  
       'slightly pale'

The meaning of the diminutive suffix is thus anti-elative. A high degree on a scale of degree that applies to a property (as expressed by the adjective) would be called elative. Given that the diminutive implies a low degree on this scale, its meaning is thus anti-elative.

As for honorifics, McCready (2019) analyses their semantics as involving a triple scale of politeness values: a scale for psychological distance, one for social distance and one for formality. We can analyse the honorifics as expressing again a low degree of distance on the scale of formality:

- (50) a. merci-ke-s  
thank.you-KE-s  
'thanks'  
b. hallo-ke-s  
hello-KE-s  
'hi'  
c. ciao-ke-s  
ciao-KE-s  
'bye'

The three empirical domains thus clearly have something in common: they all introduce a scale (size, degree, social distance) and they imply that value on the scale is low.

For a formalisation of the proposal, first consider that in Distributed Morphology, a correspondence between morphosyntax and semantics is established by proposing syntactic features that have a clear semantic definition and by decomposing syntactic properties in such features (Harbour 2008: 2-3 and De Belder to appear for discussion). Such analyses, therefore, rely on the decomposition of syntactic concepts into features that arguably have the status of semantic primitives. PERSON, for example, is decomposed into features such as [+/-Participant in Speech Event] and [+/- Author of Speech Event] rather than assuming values such as 1, 2 and 3 (see, for example, Halle 1997). NUMBER is decomposed into [singular] and [augmented], rather than assuming [singular], [dual] and [plural] as primitives (Noyer 1997, Harbour 2008). The syntactic structures in this approach are then built semantic feature per semantic feature in a strict compositional fashion. Morphological exponents are argued to realise (subsets of) precisely these features. By defining syntactic features with clear semantic content, and by proposing a clear relation between these features and the exponents, a close match is proposed between the three representations, viz. the representations at Syntax, LF and PF.

Against this backdrop, I propose that *-ke* realises the syntactic features that are responsible for the derivation of these semantic features in the tree. De Belder (2011: 122) explicitly defines the diminutive as introducing a measure function that is defined as scale, and a low value on this scale. In the case of nominal diminutives, this scale is a measure function: it assigns size to a unit. What is specific to the adverbs is the fact that the scale is understood as a degree of a quality. The honorific is conceived of as a scale of politeness. The syntactic features of the nodes can be schematised as follows:

- (51) noun.diminutive = (scale, low degree, size)  
adverb.anti-elative = (scale, low degree, intensity)  
interjection.honorific = (scale, low degree, politeness)

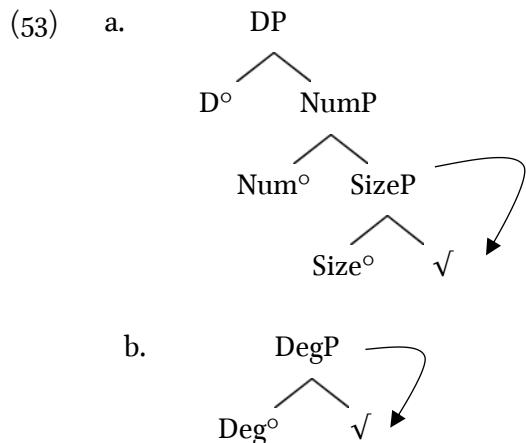
The morphemes *-ke/-tje* then realise the subset of features shared by these nodes:

- (52) /ke/  $\leftrightarrow \{[\text{scale}], [\text{low degree}]\}$   
/tje/  $\leftrightarrow \{[\text{scale}], [\text{low degree}]\}$

As such, at vocabulary insertion, these affixes can realise these three different nodes.

### 10.7.2 The insertion of the epenthetic consonant

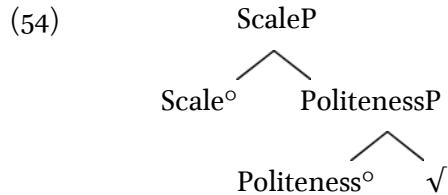
Having established what is shared by the three empirical domains, I will now discuss how the diminutives and adverbs differ from the honorifics. Consider for each domain which scale is involved in the semantics. In a diminutive the size scale is applied to what the noun refers to. For example, for *appeltje* ‘small apple’ the size scale refers to apples: a small apple is small for an apple. It is still much bigger than many other things, such as bread crumbs. In adverbs and adjectives, the degree is applied to what the embedded adjective refers to: *ziekjes* implies a low degree of illness. In other words, the scale scopes immediately over the root, and the head that introduces the scale is arguably the first functional head to merge with the root (see De Belder 2011a: chapters 3 and 4). This is shown in the trees below. The tree in (59a) is a nominal structure. The Size head refers semantically to a scale and it will be realised by the diminutive affix. Size scopes semantically over the root. For example, if the root is *apple*, a diminutive on the Size head means that the apple itself is small. The structure in (59b) is an adjectival structure. The root is assumed to become adjectival by the mere presence of a degree head, which is taken to be inherently adjectival. (The adverb would contain this structure plus an extra adverbial head). Again, the semantics of the degree scale scopes over the root itself, as indicated by the arrow. If the root is, for example, *ziek* ‘ill’, the root plus the diminutive affix together means somewhat ill.



As such, the functional head and the root are in the same Spell-out domain and they are able to condition one another.

However, for honorifics, the meaning of the root is irrelevant. To wit, *mercikes* does not refer to a low degree of gratitude, even though *merci* means ‘thank you’. The suffix *-ke* simply adds to the

informal character of the form. In other words, the scale applies to the domains of politeness, as defined by McCready (2019). More specifically, it scopes over the measure ‘formality’. I assume politeness is introduced into the structure by means of heads in the left periphery (Ross 1970; Hill 2013; Haegeman and Hill 2013; Tang 2015; Heim 2016 et al.). As pointed out before, following the semantic analysis of McCready (2019), I assume there are at least three heads involved in the cartography of politeness: social distance, psychological distance, and formality. After all, we have seen that at least social distance and formality can be set apart empirically for the Dutch honorifics. The dragged tone of *hey* implied acquaintedness (low social distance), whereas the honorific *-ke* implies a low degree of formality. We can now define a difference between diminutives and honorifics: diminutives scope over the semantics of the root, honorifics scope over the semantics of a politeness head. As such, the honorific is not the first head that merges with the root, it is merely a part of the CP-domain. It is therefore not in the same phase as the root. The functional head and the root are not spelled out together, as can be seen in the structure below. (The head politeness hosts the features ‘social distance’, ‘psychological distance’ and ‘formality’).



Note that the specific syntactic head in the C-domain has both a semantico-pragmatic effect (social distance versus formality) and a phonological effect (dragged tone versus unmarked intonation). This follows immediately from the Y-model of the language device: the syntactic structure is spelled out both to the logical form, responsible for semantics, and the phonological form. What we can observe, therefore, is interface meandering, which instantiates morphological extravagance.

This difference must be what determines the availability of the epenthetic consonant: the epenthetic consonant is only selected if the relevant functional head and the root spell out in the same phase. This can easily be formalised by formulating two separate insertion rules, one that specifies an insertion context sensitive to a root, and one that does not (more specific rules will always win from less specific rules due to the Subset Principle (see above)):

- (55) /*(s)ke/*  $\leftrightarrow \{[\text{scale}], [\text{low degree}]\}/$   
 $\quad (\{[\text{scale}], [\text{low degree}]\} \_\_ \text{Root})$   
*/ke/*  $\leftrightarrow \{[\text{scale}], [\text{low degree}]\}$

Recall that the vocabulary item which is the closest match to the syntactic structure should be inserted. When the more specific vocabulary item (i.e. the one with the richer insertion rule) is an appropriate match, it will always be a closer match, given that it has more specifications that match.

When the more specific vocabulary item is indeed an appropriate match for the syntactic structure, the diminutive will then be inserted as *-(s)ke*, leaving the presence or absence of the [s] to Phonology (cf. De Belder 2020 for optionality at Vocabulary Insertion). For the honorific, no such epenthetic consonant is available.

### 10.7.3 A remaining puzzle

We now may understand what determines the insertion of *-ke* in all syntactic contexts and we have hypothesised that a semantic distinction between diminutives and honorifics, which is reflected in very different syntactic structures, may determine the availability of the epenthetic consonant. However, what the present proposal fails to capture is the fact that *-ke* is sensitive to different Obligatory Contour Principles for diminutives and honorifics. In diminutives, only the place feature [velar] was relevant. Yet, for honorifics, it is even illicit if the root ends in a segment with low sonority. Compare the diminutive and the honorific in the following example:

- (56) a. een schaap-ke  
a sheep-DIM  
'a small sheep'
- b. koest-ke-s!  
hush-HON-s  
'Hush!'

This means that when phonology evaluates such forms, it must somehow 'see' the underlying syntax of the forms. However, much recent work in Distributed Morphology is reluctant to adopt the view that phonology sees syntax. There is a growing body of literature proposing to sever syntax from phonology when it comes to realising a structure (Bye and Svenonius 2012; Svenonius 2012; Wolf 2008; De Belder 2020 and see Booij 1998 for a similar, earlier proposal outside of Distributed Morphology). It is now often proposed that vocabulary insertion can read syntactic features and insertion contexts and that phonology only sees phonology. As such, this observation is problematic. I believe that, at this point, we cannot fully capture the observed patterns.

## 10.8 Conclusion

In this paper we have seen that the Dutch affix *-ke* and *-tje* can realise syntactic heads that are specified for the features {[scale], [low degree]}. These features occur with nouns as diminutives, in adjectives and adverbs as part of the semantics of degree, and in honorifics as part of the semantics of politeness. In this respect, we can attest *-ke/-tje* categorial unfaithfulness, which characterizes this affix as extravagant. We have further seen that the exponent *-ke* is sensitive to its phonological context. However, its phonology differs in nouns and adjectives/adverbs on the one hand and honorifics on the other. In all contexts it obeys the Obligatory Contour Principle. However, in nouns and adjectives/adverbs it cannot attach to velar consonants, in honorifics it is illicit next to velar consonants and consonants with a low degree of sonority. In nouns and adjectives/adverbs an epenthetic consonant [s] can come to the rescue. We have seen that this epenthetic consonant is a syntactic epenthetic consonant in the typology defined by Zygis (2010). I have proposed that it is restricted to nouns and adjectives/adverbs as the functional head that projects a scale that is in the same cyclic domain as the root in these syntactic contexts. This is not the case for the honorifics.

In sum, in this contribution I have pointed at an interesting empirical domain which raises problems for our understanding of the syntax-phonology interface, as it shows extravagant meandering through these interfaces. I have hinted at some pieces of the puzzle, but some

observations have not been accounted for. Most notably, the fact that the diminutive and the honorific behave differently when it comes to the Obligatory Contour Principle is a remaining puzzle.

# **11 Category-specific syllable structures in Dutch: experimental evidence**

Marijke De Belder & Esther Ruigendijk

## **Abstract**

Trommelen (1989) observed that Dutch nouns and verbs seem to differ in syllable structure, but this hypothesis has never been fleshed out, nor observationally or explanatorily. If Trommelen was right, Dutch would be one of the languages which show suprasegmental noun-verb dissimilarities (Smith 2011). In this article we pose the question whether this observation can be confirmed by means of contemporary experimental tools. This article indeed presents evidence for phonological noun-verb dissimilarities in Dutch (see also Trommelen 1989; Don and Erkelens 2006). We were able to confirm that verbal stems contain less syllables and less complex syllables than nominal stems through corpus-based computational research. In a large-scale online experiment, we also found that native speakers have intuitions on the word class of a pseudo-word on the basis of its syllable structure. Again, the intuitions are steered by the syllables' complexity and the number of syllables. Importantly, the experimental results show somewhat different patterns than the corpus research. This indicates that something steers the intuition in the linguistic competence that transcends mere awareness of statistical patterns in the lexicon. These results are discussed in the light of the Optimal Paradigms approach, which proposes an explanation for similar facts in other languages by pointing at inflection differences between verbs and nouns (McCarthy 2005; Cable 2004). Based on our findings, inflectional differences cannot explain the Dutch facts and we falsified the hypothesis that the paradigm as a whole can put pressure on the form of the stem in an experiment. As such, the study supports Bobaljik's (2006) scepticism on referring to paradigms when accounting for phonological noun-verb dissimilarities.

## **11.1 Introduction**

Trommelen (1989) observed that verbal stems in Dutch tend to be simpler than nominal stems in two respects: verbal stems tend to be monosyllabic, whereas nouns can be polysyllabic. Furthermore, the syllable structure of verbs is less complex in the sense that it tends to contain less filled time units. Indeed, speakers of Dutch have intuitions on the category of a pseudo-word. They 'feel' that the pseudo-word *pardijf* is fine as a noun -but not as a verb-, *streumel* and *preug*, on the other hand, feel more like verbs. We know that what steers this intuition is the syllable structure of the pseudowords.

Trommelen's observations are interesting, but also problematic, as she did not describe the methodology that led to these conclusions. To the best of our knowledge, these observations have also not been submitted to further research, apart from a small contribution by Don & Erkelens (2006), to which we come back below. Yet, we know from typological work that noun-verb contrasts in the suprasegmental phonology have been observed for languages with very different phonologies and morphologies. Such work raises important questions on the interface between syntax, morphology and phonology. Does the syntactic structure determine the suprasegmental phonology

of a lexical item? Does the suprasegmental phonology of a lexical item facilitate identifying the category and thus facilitate the syntactic processing of a sentence? Research on Dutch may certainly contribute to this body of work.

The aim of the present contribution is to take advantage of contemporary tools and theoretical frameworks to confirm and flesh out Trommelen's original observations, both observationally and explanatorily. More specifically, we address the following research questions. The main question is: "Is the syllabic structure of Dutch verbal stems indeed less complex than the syllabic structure of Dutch nominal stems?" The first subquestion addresses the Dutch lexicon: "Does the Dutch lexicon show phonological noun-verb dissimilarities?" It will become clear that we were indeed able to confirm such a distinction using corpus research and an artificial learner. The second subquestion questions whether this knowledge is psychologically real. In other words, is this knowledge part of the native speaker's phonology? We will show that, indeed, this held true in a large-scale experiment.

We then discuss accounts that have been formulated for similar phenomena in other languages. We present the phonological approaches in Optimality Theory and Optimal Paradigms and morphosyntactic proposals in Distributed Morphology. The most straightforward proposal to capture the facts has been formulated by McCarthy (2005) in the Optimal Paradigms framework. Discussing data from Arabic, it states that the phenomenon can be reduced to inflectional differences between nouns and verbs. The main intuition is that a word is an optimal verb if its inflectional paradigm as a whole is phonologically optimal. Hence, verbal inflection may impose important restrictions on the syllabic structure of verbs. This raises our third subquestion: can a difference in the syllabic structure of nouns and verbs be reduced to inflectional differences? Our corpus research shows that it is implausible that there is a causal relation between inflection and category-specific syllable structure. This finding is further fleshed out by the fourth and final subquestion: does the native speaker evaluate the inflectional paradigm when evaluating the syllabic structure of a verb? The question is of immediate importance to the present data set as it may account for the observations. However, we did not find any evidence that native speakers evaluate the phonology of the entire paradigm when evaluating a verb. As such, it will become clear that the verbal paradigm is not responsible for the psychological reality of the verb-noun contrast in Dutch syllable structure.

This paper is structured as follows. The next section describes background information on Dutch syllable structure and previous research on phonological noun-verb dissimilarities for this language. Sections 3 and 4 discuss what is known on this topic from a typological and theoretical point of view. Section 5 describes the computational corpus-research we carried out and section 6 describes the experiment with native speakers. Section 7 discusses whether we can adopt the insights from Optimal Paradigms to account for the phenomenon. We present a second experiment that falsifies the hypothesis that speakers evaluate entire paradigms. Section 8 is a final discussion and section 9 concludes the paper.

## **11.2 Background: The Dutch syllable and its category-specific properties**

### **11.2.1 The structure of the Dutch syllable**

The Dutch full syllable may contain an onset and a rhyme and extra-syllabic positions for a word's final rhyme. The onset has zero to two filled time units and the rhyme contains at least two and at

most three filled time-units. Both before the onset and after the rhyme zero to two extra-syllabic time units can be found. Extra-syllabic positions are always occupied by either [s] or [t]. Lax vowels fill one time unit, tense vowels and diphthongs fill two time units (Trommelen 1983, Kager & Zonneveld 1986)<sup>182</sup>. The following table (Table 1) summarises and illustrates these facts.

Table 1: Time units in the Dutch syllable

extra-syllabic	onset		rhyme			extra-syllabic		example
x	x	x	x	x	x	x	x	
		z	e					zee 'sea'
s	x	r	u		f			schroef 'screw'
s	x	r	i	f	t			schrift 'notebook'
	b	r	a	n	d			brand 'fire'
	g	r	a		n			graan 'grain'
		t	o		r	n		toorn 'wrath'
		h	ɛ	r	f	s	t	herfst 'autumn'
			a		p			aap 'monkey'

For reduced syllables, i.e. schwa-syllables, a single filled time-unit position in the rhyme suffices. For monosyllabic words this occurs for functional vocabulary, as in the definite article *de* /də/ 'the'. Reduced syllables may also occur as one of the syllables of a polysyllabic lexical word, as in the examples *beamən* /bəamən/ 'to confirm' and *lente* /ləntə/ 'spring'.

### 11.2.2 Trommelen's (1989) observations

Trommelen (1989) observed that the syllable structure of a morphologically simplex Dutch word may be indicative of its category. She argues that verbs have the most simple syllable structures, adjectives show somewhat more syllabic possibilities and nouns show the least restrictions qua syllable structures. More specifically, she argued that the possible number of the syllable's final consonants and the number of syllables may vary across categories. She considered the complexity of the onset to be irrelevant for cross-categorial distinctions. (It will become clear in section 5.4.1 that a more nuanced view on the category-specific role of the onset is warranted.) Her article focused on word-final syllables, for good reason, as the rhyme structure of word-internal syllables is mostly limited to two positions anyway in all Dutch words.

---

<sup>182</sup> We abstract away from the theoretical question whether these positions belong to the final syllable or to the phonological word as a whole. The internal structure of the rhyme is not relevant here either and we ignore it for ease of exposition.

In what follows, we will ignore Trommelen's observations on adjectives, as they fall out of the scope of the present work. Her findings on monosyllabic words are summarised in table 2 below. Note that for verbs, we are discussing the stems, rather than the infinitives. The column header describes the consonantal cluster at the end of the syllable: no consonant, a sonorant, an obstruent, paired combinations of sonorants and obstruents or three consonants. Dutch syllables with lax vowels and a final consonant cluster of four consonants do exist in Dutch (e.g. *herfst* 'autumn'), but they are not included in Trommelen's discussion. The row headers describe the nature of the vowel: a lax monophthong, a tense monophthong, a diphthong or a schwa. The cells contain a real-word example that illustrates the combination of the vowel and the consonant cluster of the cell. The color of the cell shows whether nominal and verbal stems can occupy these cells. White cells can be occupied by both categories. The black cells can be occupied by neither of them, either because the syllable structure is excluded from Dutch or because it is excluded for lexical categories in case of schwa syllables (which do exist in the functional vocabulary, as illustrated by the definite articles *de* and *het* and the indefinite article *een*.) The light grey cells are cells that can only be occupied by nouns or by verbs that are converted from nouns, according to Trommelen. The dark grey cells are cells which exist marginally, according to Trommelen: less than five examples could be found and these examples are most often nouns or verbs that are converted from nouns.

Table 2: Trommelen's observations on Dutch monosyllabic words

		final consonantal cluster						
		zero	son.	obstr.	son. + obstr.	obstr. + obstr.	son.+ son.	CCC
vowel of the syllable	lax monoph.	sjwa 'schwa'	ram 'ram'	rat 'rat'	kalf 'calf'	gips 'plaster'	worm 'wurm'	worst 'sausage'
	tense monoph.	zee 'sea'	riem 'belt'	noot 'nut'	moord 'murder'	feest 'party'	hoorn 'horn'	koorts 'fever'
	diphthong	kei 'boulder'	duim 'thumb'	meid 'maid'	eind 'end'	beits 'stain'		
	schwa							

monoph. = monophthong, zero = no consonant, son. = sonorant, obstr. = obstruent, CCC = 3 consonants

Trommelen (1989) notes that there are exceptions to her observations. For example, there are verbs that contain a diphthong and that end in a sonorant plus obstruent, such as *peins* 'reflect' and *deins* 'recoil'. Indeed, Trommelen's observations should be understood as strong tendencies rather than as absolute generalisations. This point will be a central tenet in all analyses below. Empirically, this is a domain that should be expressed in terms of *probabilities* rather than in terms of *generalisations*.

In the discussion of monosyllabic words, Trommelen included bisyllabic words of which the second vowel is a schwa, such as *adem* 'breath' and *fluister* 'whisper' due to her theoretical conviction that these words are in fact monosyllabic (see Kager & Zonneveld 1985-1986; Trommelen

1989): the schwa plus sonorant was analysed as occupying extra-syllabic positions that belonged to the initial syllable. Structurally, the entire word was therefore analysed as one single syllable. For ease of exposition, we will not follow Trommelen in this somewhat idiosyncratic choice and we include these words in the discussion of the polysyllabic words.

Trommelen's observations on polysyllabic words are summarised in table 3. The table has the same legend as the previous one. The vowels and the consonantal clusters described by the header describe the nature of the words' final syllable.<sup>183</sup> The same color codes apply as in table 2.

Table 3: Trommelen's observations on Dutch polysyllabic words

		final consonantal cluster						
		zero	son.	obstr.	son. + obstr.	obstr. + obstr.	son.+son.	CCC
vowel of the syllable	lax		bizon 'bison'	barok 'baroque'	concert 'concert'	orkest 'orchestra'	alarm 'alarm'	
	monoph.							
	tense	villa 'villa'	banaan 'banana'	tomaat 'tomato'	akkoord 'agreement'	langoest 'spiny lobster'	lantaarn 'lantern'	
	monoph.							
	diphthong	abdij 'abbey'	fontein 'fountain'	fornuis 'stove'	accijns 'tax'			
	schwa	lente 'spring'	adem 'breath'	monnik 'monk'	mosterd 'mustard'			
	schwa +l/r		fluister 'whisper'					

monoph. = monophthong, zero = no consonant, son. = sonorant, obstr. = obstruent, CCC = 3 consonants

Trommelen's observations on polysyllabic words can be summarised easily: they are fine as nouns, but they do not occur as verbal stems unless they result from denominal conversion. Again, these observations are probabilistic rather than absolute. There certainly are exceptions, such as *neurie* 'hum' and *bakkelei* 'quarrel'. One cell is a clear exception to the rule: if the second syllable is a schwa plus an /r/ or an /l/, a bisyllabic structure may be verbal, as in *fluister* 'whisper', *stotter* 'stutter' and *stamel* 'stammer'. A schwa plus a sonorant other than an /r/ or an /l/ does not fall under this exception. Trommelen hesitates for two cells: the polysyllabic structures that end in a diphthong and those that end in a lax vowel plus an obstruent. For both cells a handful of verbal examples can be found, such as *bakkelei* 'quarrel', *plavei* 'pave', *ravot* 'romp', *slabak* 'slack'.

Trommelen concludes that it is possible to deduce categorial information from the syllabic structure of a morphologically simplex stem. The possibilities of a verbal rhyme structure are a

<sup>183</sup> Trommelen provides no data for the cell 'schwa + sonorant + obstruent'. I assume this was an error of oversight on her part. This type of noun is indeed marginal, but relevant examples exist: *wereld* 'world', *arend* 'eagle', *mosterd* 'mustard'.

subset of the nominal ones. Furthermore, the inventory of morphologically simplex polysyllabic verbal stems is negligible.

### 11.2.3 A first experiment: Don and Erkelens (2006)

Don and Erkelens (2006) designed a small-scale experiment to test Trommelen's (1989) observations. More specifically, they aimed to test the hypotheses that the following structures are less accepted as verbal stems: polysyllabic structures with full vowels, words ending in a monophthong, words ending in a consonantal cluster that makes the syllable more than superheavy, words ending in a schwa and words ending in a schwa followed by an /m/. These hypotheses correspond to Trommelen's conclusions.

The experiment consisted of 20 Dutch pseudowords, 11 of which were designed along the lines of Trommelen's (1989) conclusions to be recognised as nominal stems and 9 of which were designed to be ambiguous between verbhood and nounhood. The task was printed on paper. There were three columns: the first column contained the pseudowords, the second column had the word 'verb' in the header and the third column had the word 'noun' in the header. Participants were asked to put for each pseudoword an X in the column of their preference. The words were given without any context, to avoid semantic or syntactic influences. The task was preceded by a short explanation in which it was made clear what 'verbal stem' and 'nominal stem' meant. 28 adult, monolingually raised participants completed the task. The results are given in figure 1:

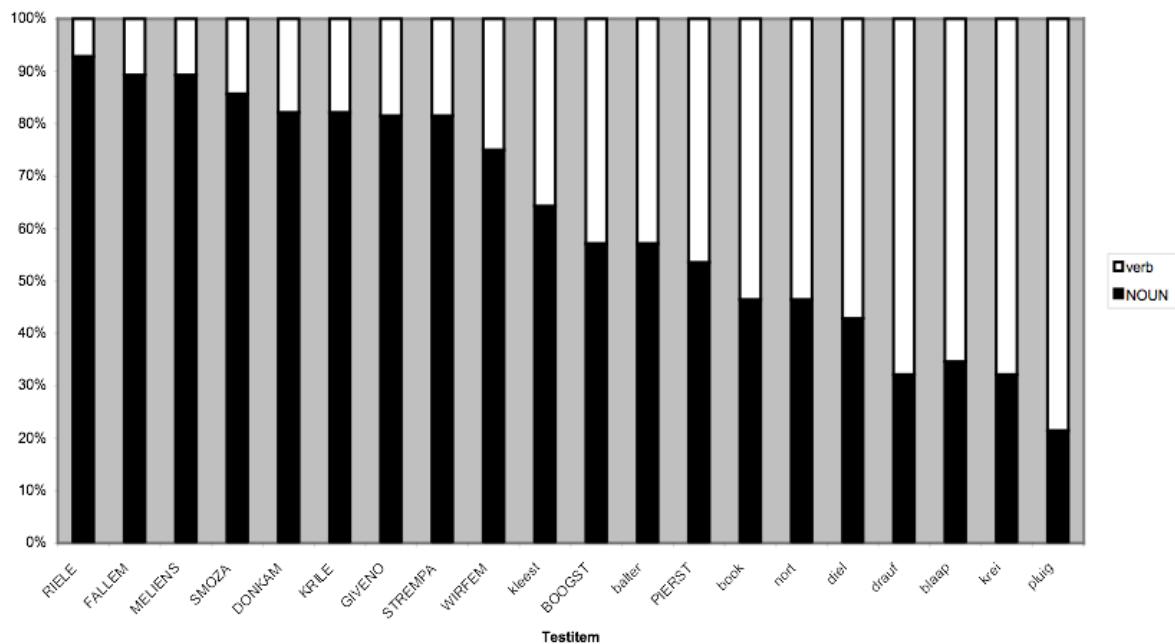


Figure 1: Results from Don and Erkelens (2006), taken from Don and Erkelens (2006:49)

Don and Erkelens observe that Trommelen's (1986) observations are confirmed by this experiment, with the exception of the observation that syllables with extra-syllabic endings appear to be perceived less as nouns than expected on the basis of Trommelen's study (see the results for *pierst* and *boogst*). It will become clear later in this article that we can confirm this observation: extrasyllabic material is a constraint on the syllable structure of verbs, but a less severe one than polysyllability. Note that Don and Erkelens did not have a monosyllabic test item that ended in a

monophthong. Importantly, Don and Erkelens (2006) conclude that knowledge on category-specific syllable structures is part of the native speaker's language competence.

#### **11.2.4 Interim conclusion**

Trommelen (1989) observed that the syllabic structures of morphologically simplex Dutch verbs show restrictions that nouns do not. The inventory of polysyllabic verbal stems is negligible and the possibilities of a verbal syllable structure are a subset of the nominal ones. The results of the experiment of Don and Erkelens (2006) were in line with most of these observations and indicated that the category-specific syllable structures are part of the native speaker's phonology.

The work by Trommelen on the one hand and Don and Erkelens on the other revealed important novel insights on Dutch word structure. However, both articles show important methodological caveats. Trommelen (1989) undeniably performed a thorough study, but she did not describe her methodology, which makes it difficult to evaluate the validity of her observations. Don and Erkelens' (2006) experiment shows strong results, but they themselves already acknowledge some caveats. It was a small-scale experiment, with only one test item per condition. The design also required some linguistic meta-awareness of the participants as they were explicitly asked to sort the pseudowords as nouns or verbs. The test items were not controlled for phonological neighbourhoods either. As such, it cannot be excluded that judgements were based on word associations. In sections 5 and 6 we remedy these caveats by repeating the two studies. We repeated Trommelen's study using a corpus and an artificial learner. We repeated the experiment by Don and Erkelens (2006) on a large-scale basis after having modified the design of the experiment. It will become clear that we were able to verify most of the findings of these earlier works. However, before discussing these findings, we first discuss typological and cross-linguistic work on category-specific suprasegmental phonologies.

### **11.3 Background: Typological insights on category-specific suprasegmental phonologies**

Dutch is certainly not the only language in which different phonological rules or constraints seem to apply to different categories. Smith (2011) presents a typological overview, which is summarised in this section. It is not uncommon for a category in a language to show what she calls phonological privilege: the ability to support a greater array of phonological contrasts. These contrasts are usually suprasegmental or prosodic and typically include stress, accent, tone and syllabic positions. Privilege at the segmental level that cannot be reduced to a difference in moraic positions has rarely been attested and the few relevant examples are disputed as alternative analyses are often available.

The generalisation seems to be that noun privilege is the norm (see also Smith 1997). Noun privilege has been widely attested across different language families. For example, in Spanish and Hebrew, stress location is contrastive for nouns, but predictable for verbs (Harris 1983; Garrett 1996). Pitch accent location shows a similar pattern in Tokyo Japanese (McCawley 1968). The same holds true for tone in Mono (Olson 2005). In Hebrew, verbs must fit into disyllabic templates, nouns may be templatic, but they do not need to be (Bat-El 1994:577-578). In Mbabaram long vowels are only attested in nouns, which is seen as a moraic privilege rather than a segmental one.

For several languages, it has been observed that nouns are the only category that occur as free morphemes (e.g. Japanese) or that have to observe word minimality requirements. For example, in

Chuukese (Muller 1999) nouns have to be at least bimoraic and when they are not because of a truncation process in the language, the nouns are augmented by means of vowel lengthening. Verbs do not show such augmentation. This suggests that noun privilege does not only involve a greater array of phonological contrast, but a general tendency to be heavier in the sense that they allow or even require more filled moraic positions. In an article more recent than Smith (2011), Holmberg and Wang (2019) observe that in Traditional Xining Chinese (and other dialects spoken in North West China), free nouns are always reduplicated, a phenomenon for which they present a morphosyntactic account.

Smith (2011) presents one solid example of a language with verb privilege: in Ewe tone may be neutralised in certain contexts for nouns, but not for verbs (Ansre 1961). She lists more potential examples of verb privilege, but warns that it is unclear for these examples which category in fact shows the most marked behaviour. In short, noun privilege is widely attested, verb privilege is extremely rare or at least disputed.

According to Smith (2011), adjectives may pattern both with nouns or with verbs or they may be an intermediate category. She concludes that for languages with suprasegmental phonological categorial contrasts a N > A > V ranking for phonological privilege is a (near-)universal. At this point, it is worth pointing out that Trommelen's (1989) observations on Dutch are in line with Smith's typological observations. Trommelen argued that nouns show more possibilities than verbs and she argued that adjectives fall in-between nouns and verbs in Dutch. Also the prosodic nature of the contrast fits within Smith's observations. The next section discusses some accounts that have been presented to capture the observations.

## 11.4 Background: A phonological phenomenon or a side-effect of interfaces?

### 11.4.1 Introduction

Smith (2011) briefly discusses some accounts that have been presented to capture cross-categorial phonological distinctions. The main question is whether the phenomenon is phonological in nature or rather a side-effect of morphology or syntax. For example, for some languages, the noun-verb distinction seems to be reducible to a free-bound distinction in morphology which maps neatly onto a noun-verb contrast. Other proposals to capture the phenomenon refer to inflectional differences, prosodic differences or morpheme-specific differences. Smith warns for certain problems with such proposals and she points out that, even though it is clear that such factors certainly can account for some data in some languages, it is also clear that the field has failed so far to reduce the phenomenon to such side-effects. Below, I discuss some of the current approaches in more depth. In the next section, I discuss proposals that approach the phenomenon essentially as a phonological issue, section 4.3 presents approaches that take the angle of the phonology-morphology or phonology-syntax interfaces, section 4.4 discusses the Optimal Paradigms approach.

### 11.4.2 Interface constraints and multiple phonologies

Optimality Theory (OT) is an approach to phonology in which phonological rules are formulated as ranked constraints. These constraints serve to define the most optimal surface form on the basis of a given underlying phonological form. The constraints are given by Universal Grammar, but their ranking is language-specific.

Within Optimality Theory, category-specific phonological constraints have been captured by positing interface constraints and by means of co-phonologies (see Anttila 2002:1-3 for discussion and see references therein). Interface constraints are specific to certain classes. They have, for example, been used to distinguish between the phonological behaviour of bound versus free morphemes or between various lexical strata. We have just seen that Smith (2011) points out that attempts to reduce a noun-verb distinction to a side-effect of other phenomena, such as the free versus bound distinction and differences in prosodic structure is bound to fail cross-linguistically.

Reducing the noun-verb distinction to other morphological distinctions may not be straightforward. However, extending the *logic* of specific constraints for specific lexical items, one can postulate noun-specific or verb-specific faithfulness or markedness constraints (Smith 2011). For example, Smith proposes Noun Faithfulness constraints. These are phonological constraints which are ‘indexed’ to apply only to nouns. Needless to say, such a proposal raises important questions on how to organise the interfaces between syntax, the lexicon and phonology.

#### 11.4.3 Morphosyntactic accounts

A straightforward way to capture category-specific phonologies is to relate them to morphosyntactic differences. After all, we know that nouns and verbs have their own specific functional spine. The distinction comes for free: at syntax the categories differ. One then just has to find a way to relate the (morpho)phonological make up of either the verb or the noun to a morphosyntactic requirement to capture noun-verb distinctions. Several of such accounts can be found in the framework of Distributed Morphology (Halle and Marantz 1993, 1994; Harley and Noyer 1999). Below we briefly mention three such approaches: one that singles out nouns as the category with a specific demand, one that singles out verbs and one that reduces the difference to a distinction in the size of the spell-out phase. We mention these approaches for good measure. It is on our research agenda to flesh out these ideas experimentally, but we do not tackle this within this article.

Bobaljik (1998) discusses noun-verb contrasts from Itelmen. In Itelmen, the rules for schwa insertion differ for the two categories. Bobaljik argues that Itelmen has schwa insertion for syllabification. For nouns, schwa insertion depends on the phonology of the root and on whether a following suffix is consonant or vowel initial. For verbs, schwa insertion depends on the phonology of the root alone and it will always take place in the relevant context regardless the initial segment of a following suffix. For the first phase of a cyclic derivation only the root is visible. Bobaljik argues that at this point of the derivation the rule for schwa insertion applies to a verbal derivation: the rule of schwa insertion applies cyclically in verbs. For nouns, the rule applies non-cyclically and thus takes into account the entire structure, including the suffix. As such, the approach derives the noun-verb contrast directly from the syntax-PF interface. The question remains, of course, why a rule would apply cyclically to verbs and non-cyclically to nouns.

As already mentioned above, Holmberg & Wang (2019) describe that in traditional Xining Chinese, free nouns are always reduplicated. Working within Distributed Morphology, they argue that the structure of a noun in syntax involves an acategorial root which is nominalised by a functional node called little n. It is this nominalising head which is realised, as they argue, by the reduplicated root. Within Distributed Morphology, a content word will always consist of minimally two nodes, i.e. the root and the little head which assigns a category to the root. What is special about nouns in Xining Chinese, according to Holmberg and Wang, is that a noun must minimally consist of two *overtly realised* nodes (see also Smith 2011 for various other languages and Goodenough and

Sugita 1980 on Trukese, as mentioned in Cable 2004). Of course, the question is why *nouns* specifically require two pronounced roots in Xining Chinese (and perhaps other languages) and other categories do not.

Rather than arguing why nouns need to be so big, one could equally find a reason why verbs need to be so small in comparison. Arad (2005) takes this route. In Hebrew, verbs need to be templatic, which means that they need to consist of precisely defined configurations of consonantal and vocalic slots. Nouns may be templatic, but they do not need to be so. Arad argues that the consonants in the verb realise the root and that the vowels realise the functional node Voice, which defines the verb as active or passive. It is an interesting idea to propose that verbs need to obey specific suprasegmental restrictions because of the need of a functional node to be realised. It is immediately clear why the restriction is specific to verbs: nouns obviously do not have a Voice projection.

It would certainly be interesting to explore the idea that verbs have to particularly small or nouns have to particularly large enough for specific syntactic reasons and we hope to address such ideas in more detail in future work.

#### 11.4.4 Optimal Paradigms

Optimal Paradigms is an Optimality Theory-based approach in which candidates for evaluation are entire inflectional paradigms. Inflectional paradigms contain all the words based on a single lexeme. Violation marks incurred by each paradigm member are added to the total sum of the paradigm for evaluation (McCarthy 2005).

In section 4.2 we have seen that many OT approaches will ascribe noun-verb contrasts in phonology to category-specific constraints or to cophonologies. In other words, the contrast is seen as a *phonological* issue. Optimal Paradigms, however, follows a different road. They see the contrast as an epiphenomenon of inflectional differences, which are evaluated at Phonology (McCarthy 2005; Cable 2004).

It is argued that all members of a paradigm stand in a relation to one another. Inflected forms are thus psychologically linked to the other inflected forms of the same paradigm, rather than to a base. Out of faithfulness concerns, it is argued that it is optimal for members of the same paradigm to be uniform. Such uniformity may be a highly ranked constraint. A consequence is that an inflectional requirement that applies to one member of the paradigm may impose a restriction on the base as a whole. If an inflectional affix imposes a certain restriction on a base for one member and if uniformity of the base is a highly ranked constraint, the restriction will limit the base for all members of the paradigm. Imagine, for example, that a specific verbal suffix can only apply to vowel-final stems and that base-uniformity is important. The constraint may then be responsible for the selection of base that has been transformed such that it becomes vowel-final and, moreover, this transformed version of the base may become uniform for all members of the paradigm. In such a language, the verb may appear to be vowel-final throughout the paradigm. In fact, the suffix may be responsible for the fact that *all* verbs are *always* vowel-final in the language. As such, category-specific phonological requirements may emerge.

Phonology is thus not sensitive to categorial distinctions, it is rather sensitive to the requirements imposed by inflectional affixes, which happen to differ across categories. McCarthy (2005) presents a detailed analysis of noun-verb contrasts in Classic Arabic, Cable (2004) mentions tonal noun-verb dissimilarities from Shona, different moraic minimal requirements for verbs and

nouns in Trukese, Arabic verbal templates as discussed by Bobaljik 2002 and McCarthy 2005 and Bobaljik's (1998) data from Itelmen.

Cable (2004) hypothesises that Optimal Paradigms is responsible for all attested category-specific phonological patterns. The attractive part of that hypothesis is that it has strong explanatory power. But does it hold true descriptively? Cable himself indicates that he realises the hypothesis entails a bold claim empirically: all involved languages should show relevant inflectional patterns that can explain the facts. In section 7.1 we will discuss whether *in principle* it can hold for the Dutch facts. We will show that the reasoning could indeed account for a part of the Dutch facts. Optimal Paradigms also involves a bold claim psycholinguistically. The proposal would entail that members of paradigms are indeed psychologically related and that candidates to compute optimal phonological forms are entire paradigm sets rather than individual forms. Recall that it is hypothesised that violation marks incurred by each paradigm member are added to the total sum of the paradigm for evaluation. In sections 7.2-7.4 we present an experiment to evaluate the plausibility of this part of the theory. It will become clear that we did not find any evidence in favour of an evaluation of an entire paradigm. As such, the evidence does not support the Optimal Paradigm approach and complement the theoretical objections against Optimal Paradigms raised in Bobaljik (2006).

#### **11.4.5 Conclusion**

Accounts for phonological noun-verb dissimilarities go in one of the following two directions: either the issue is seen as a phonological phenomenon or it is seen as a by-product of the way phonology interfaces with morphology or syntax. Interface approaches then argue that nouns need to be big enough for a specific morphosyntactic reason, that verbs need to be small enough for a certain morphosyntactic reason, that syntactic cycles of verbs and nouns are simply derived differently or that either verbs or nouns need to fulfil a specific requirement imposed by inflection. The Optimal Paradigm approach is a detailed account that points at inflection. It is promising in the sense that it intuitively makes sense as an explanation: what can be inflected well as a verb, makes a good verb, what inflects well as a noun makes a good noun in a language. It is also attractive that the theory is precise enough such that it can be verified or falsified empirically. This is interesting, as it makes particular strong claims on what we may find empirically. It entails strong claims on the organisation of the human's linguistic competence: it assumes members within paradigm sets are strongly connected and evaluated together.

### **11.5 Computational corpus research: confirming Trommelen's observations**

#### **11.5.1 Introduction**

As was mentioned in the introduction, we will address various subquestions in the present contribution. We aimed to establish whether Trommelen's observations on the prosodic noun-verb contrast in Dutch hold true, whether this contrast is present in the linguistic competence of the native speaker and whether the contrast can be ascribed to inflectional differences. This section addresses the first subquestion. We have verified the validity of Trommelen's observations using corpus data and the Phonotactic Learner, a software device that artificially learns a lexicon and can report back on prosodic restrictions. We used this method to answer the following two questions.

The first question is whether the syllable structure of a monosyllabic nominal stem differs from the syllable structure of a monosyllabic verbal stem. This answer aims to research contrasts in the onset, the rhyme and the extra-syllabic positions of the syllable. We chose to use the monosyllabic part of the nominal and verbal vocabulary to have two lists that are maximally comparable. After all, if the syllable structure differs across categories, this difference should already be visible using monosyllabic lemmas. As such, we can guarantee that differences in syllable structure cannot be reduced to a more general difference in number of syllables. The second question is whether verbs indeed contain less syllables and, more specifically, less full syllables (i.e. syllables with a vowel other than the schwa) than nouns. These questions go back to Trommelen's (1989) initial work, with the difference that we did not exclude the syllable's onset from the research. In this section we describe our methodology and the results of this study. It will become clear that we were able to confirm Trommelen's claims and that we have even found a small contrast between nominal and verbal onsets.

### 11.5.2 The Maximum Entropy Model of Phonotactics and Phonotactic Learning

The corpus-based computational part of this project borrows a tool that has been developed within a model called 'Maximum Entropy Model of Phonotactics and Phonotactic Learning' (Hayes & Wilson 2007), which is embedded in the framework of Harmonic Grammar (Legendre, Miyata, & Smolensky, 1990; Smolensky & Legendre, 2006), which is a framework embedded in the more general context of Optimality Theory.

In Optimality Theory's usual approach (Prince and Smolensky 1993), the constraints are ranked: a form that violates a higher ranked phonological constraint is phonologically less optimal than a form that violates a lower ranked constraint. An alternative approach is Harmonic Grammar (HG) (Legendre, Miyata, & Smolensky, 1990; Smolensky & Legendre, 2006), in which the strength of a constraint is defined by assigning it a certain weight. Rather than strictly deriving the most optimal surface form, as in OT, HG defines probabilities per form. The advantage of this latter approach is that weights can be used to estimate the probability of forms numerically. As such, they are compatible with statistical and experimental approaches (see Pater 2009 for discussion). This approach is fully developed in a model called 'Maximum Entropy Model of Phonotactics and Phonotactic Learning' (MEMPPL) (Hayes & Wilson 2007), which defines a framework and a methodology to estimate the probability of a form. Combining the idea of defining category-specific phonologies and the idea of working with weighted constraints, allows us to define probable and improbable Dutch nouns and verbs.

Within MEMPPL a software tool is available, called the Phonotactic Learner. It is an artificial learner that can measure the weights of phonotactic constraints (i.e. constraints on possible combinations of consonants and vowels in the syllable). The software is given a representative lexicon as its input, i.e. a list of lemmas and their relative frequencies. It is also fed a feature chart that describes the phonemes of the language and their distinctive features. The program then can compute a list of phonotactic constraints and their relative weights or it can measure a weight for a constraint formulated by the researcher. Note that the software takes the frequencies of the form into account. As such, it can reach results one cannot reach by simply manually counting phonotactic combinations in the corpus.

The software assigns a weight to each constraint. This value reflects the probability of the phonotactic combinations in the lexicon. As such, the results can function as gradient measures. For example, we know that for Dutch the weight for a final voiced obstruent is actually an absolute

constraint (due to final devoicing). The constraint will then be assigned a weight by the software that is very high. To illustrate the weights assigned by the program, we have measured precisely this constraint for Dutch nouns and its weight had the value 9,178. The weight for a fully unmarked, licit phonotactic combination will be zero. (Zero means ‘no constraint’.) The weight for rare and marked, but existing phonotactic combinations will have a value somewhere in between. For example, the weight for the phonotactic combination of an empty onset in front of the vowel /œy/ for Dutch nouns has the value 6,302. This reflects the fact that this phonotactic combination is unlikely, but possible in Dutch. To get a feel of what this means: in our corpus list of nouns, this phonotactic combination occurs just once, for the Dutch noun *uil* ‘owl’, which has a frequency value of 149 in our corpus, which is considerably lower than the average frequency of 1068 for nouns (see below, table 4).

The probability-based nature of the tool is ideally suited for this empirical domain. For example, the claim that a verbal stem cannot be syllabically complex is obviously false. Trommelen (1989) herself already pointed at verbs such as *neurie* ‘hum’ and *bakkelei* ‘quarrel’. However, the claim that verbs are less likely than nouns to have a complex syllabic structure might very well be true and a model that allows us to express this contrast numerically is an ideal tool to describe the data.

The fact that we adopt the tool to describe the data, does not entail that we adopt the theoretical approach to noun-verb contrasts from OT. One can absolutely use the mathematical and computational approach to calculate the probability of the phonotactic form of a stem without accepting all premises of Optimality Theory. The results of the software are interpretable and usable, regardless whether one accepts the theoretical claim that nouns and verbs are evaluated using different constraints, as has been proposed within Optimality Theory (see section 4.2).

### 11.5.3 Methodology

We used data from the SUBTLEX-NL corpus (Keuleers, Brysbaert & New 2010). The corpus contains lemmas and frequency measures based on film and television subtitles. We selected this corpus because it is taken to be representative of the lexicon of the language-user as the corpus is based on non-edited text. Furthermore, the frequencies of subtitles perform very well as indicators in reaction time studies (New, Brysbaert, Veronis and Pallier 2007; Brysbaert and New 2009) as they are an approximation of everyday word use. It is also a relatively recent corpus. For our specific purposes it was ideal that this corpus contains part-of-speech tagging and contains frequency measures per lemma for a specific part-of-speech.

We exported all words that were tagged as nouns or verbs in the SUBTLEX-NL corpus (Keuleers, Brysbaert & New 2010), according to the rules dominant\_pos = WW (“the dominant part of speech for this lemma is verb”) for verbs and dominant\_pos = N (“the dominant part-of-speech for this lemma is noun”) for nouns, together with a frequency count of that lemma when used as that part-of-speech (“dom\_pos\_lemma\_freq”). The frequency measure is a measure of the lemma and not of a specific word-form, i.e. it sums all various inflected forms. It is category-specific, though. For example, the noun *douche* ‘shower’ has a lemma frequency of 1040, whereas the verb *douchen* ‘to shower’ has a lemma frequency of 813.

We were only interested in morphologically simplex forms, as morphologically complex forms contain more morphemes and more syllables and as such interfere with the main question whether nouns and verbs *an sich* have a different syllable structure. So, from the exhaustive lists of verbs and nouns we manually selected all morphologically simplex words. We included all words that could be considered members of conversion pairs. For example, the noun *douche* ‘shower’ and the verb

*douche* ‘to shower’ were not deleted. We included these as noun-verb conversion is productive in Dutch (Don 1993) and excluding these data would result in a deletion of a substantial part of the vocabulary. We did not exclude data that are only formally complex either. For example, a noun such as *potlood* ‘pencil’ was not excluded. It is formally complex as it looks like a compound. Its syllables are reminiscent of the morphemes *pot* ‘pot’ and *lood* ‘lead’ and its stress pattern is the stress pattern of a Dutch compound. Yet, it is generally not considered to be truly morphologically complex as the semantics of its parts do not seem to contribute transparently to its meaning as a whole. In the same vein, words such as *beam* ‘to confirm’ and *begin* ‘to start’ were included in the list as well. Formally, one might recognise the prefix *be-*, but *aam* and *gin* are not recognisable as relevant stems in these words. Therefore, such words are generally considered to be only formally complex.<sup>184</sup> By including members of conversion pairs and formally complex words we aimed not to transform the data by means of an artificial manipulation of the lexicon.

We then fed these remaining lemmas to the Dutch Clearpond<sup>185</sup> corpus, i.e. Dutchpond (Marian, Bartolotti, Chabal and Shook 2012) in order to obtain phonological transcriptions. For the lemmas for which no phonological transcription was available in Clearpond, we completed the transcriptions manually. In case of doubt, we consulted the transcription dictionary of Heemskerk and Zonneveld (2000). We changed the transcriptions such that they only transcribed the stems of the words. For the nouns this transformation is null in Dutch: the singular form of a morphologically simplex noun coincides with the stem. For the verbs this meant that we deleted the infinitival ending /ən/. For verbal stems ending in a voiced obstruent we replaced the voiced obstruent by the corresponding voiceless obstruent. We did this to unify the lists of nouns and verbs. Nouns in Clearpond are transcribed with the effect of Dutch final devoicing. For example, the transcription of *hond* ‘dog’ will contain a [t] rather than a [d]. This change did not affect our research, but we mention it for good measure.

We then replaced the symbols of the phonological transcriptions into symbols that are compatible with the Phonotactic Learner. We also tagged the stems for the number of syllables, so we would be able to isolate monosyllabic stems. Finally, we deleted all stems that contain foreign vowels (see below), as we did not define foreign vowels in the feature chart we fed to the software (see below) and as we are not interested in loanword phonology in this project. This resulted in four lists: phonological transcriptions and lemma frequencies of all verbal stems, all nominal stems, all monosyllabic verbal stems and all monosyllabic nominal stems. The number of lemmas in these lists and their summed frequencies are summarised in the following table:

---

<sup>184</sup> Admittedly, it is not always clear whether a word is formally or morphologically complex and we guess that various words may even be judged differently depending on the individual’s speaker lexical knowledge. However, we estimated that decisions regarding the few cases which would be debatable would not affect the results as a whole given the size of the corpus.

<sup>185</sup> Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities.

Table 4: Number of lemmas and their frequencies

	number of lemmas	summed frequencies	average frequency per lemma
V	1402	7307312	5212,06
V, monosyllabic	872	6769119	7762,75
N	2939	3137969	1067,7
N, monosyllabic	1218	1890901	1552,46

V = verb, N  
= noun

We would like to note that our corpus is big enough and our frequency measures are high enough to yield reliable results for the Phonotactic Learner. The Phonotactic Learner requires at least a summed frequency of 3000 or more for accuracy (Capodieci, Hayes & Wilson 2009:8), a measure which we exceeded by far.

Even though this table merely summarises the data we worked with for methodological purposes, the reader will of course notice that the numbers already indicate something about the status of the Dutch lexicon. There are more morphologically simplex nouns than there are morphologically simplex verbs and this goes hand in hand with a higher frequency per individual verb. The average frequency per lemma indicates that monosyllabic lemmas are used more frequently than the polysyllabic lemmas. Finally, 62% of the verbal stems are monosyllabic against 41% for nominal stems. The reader should keep in mind at this point that the polysyllabic data include both lemmas with several full syllables as combinations of a single full syllable and a schwa syllable. In fact, of the polysyllabic verbs 85% contain at least one schwa, whereas only 41% of the polysyllabic nouns do. We would therefore like to warn the reader against hasty conclusions based on this table: polysyllabicity is probably not a unified phenomenon across categories.

For the Phonotactic Learner we manually prepared a chart of all Dutch phonemes and their distinctive features in a feature chart. We did not include any foreign (i.e. French or English) vowels.. We also manually wrote a grammar for the program containing the constraints to be tested. These constraints systematically correspond to the cells of table 2. We thus formulated constraints such as, ‘diphthong plus obstruent plus word boundary’, ‘lax monophthong plus sonorant plus sonorant’, etc. Note that we use the word ‘constraint’ here not in the sense of ‘restriction’, but in the context of the Phonotactic Learner. Constraints may turn out to be null constraints, indicating there is no restriction. We ran the program four times: we tested the constraints for the monosyllabic verbal stems, for the monosyllabic nominal stems, for all verbal stems and for all nominal stems. This allowed us to compare the syllable structure of monosyllabic nouns with the syllable structure of monosyllabic verbs, to study cross-categorial differences in the syllable’s complexity. It also allowed us to compare the corpus’ complete inventory of morphologically simplex verbs to the complete inventory of morphologically simplex nouns to study differences in polysyllabicity.

## 11.5.4 Results

In what follows we will first describe the results regarding contrasts in syllable structure for monosyllabic stems and then we discuss the question whether the number of syllables also differs per category.

### *n.5.4.1 The syllable structure of monosyllabic stems*

For monosyllabic stems, we looked at differences between verbs and nouns for the rhyme (plus final extra-segmental positions) and at the onset (plus initial extra-syllabic positions). We first discuss the rhyme. The results for the rhymes of monosyllabic nouns and verbs are summarised in the following table (Table 5). The left column describes the nature of the syllable's vowel, the rows describe the nature of the syllable's final consonants: no consonant, a sonorant, an obstruent or combinations thereof. The N-rows reflect the weights of the restrictions for nominal stems, the V-rows do the same for the verbal stems. We also give the difference between nouns and verbs as to allow for a quick comparison.

Recall that the Phonotactic Learner measures markedness constraints. A zero corresponds to no restriction, a higher number corresponds to a restriction; the higher the number, the more severe the restriction. Of course, the higher the numerical difference between the nouns and the verbs for a given cell is, the more the syllable structure for verbs and noun differ in the lexicon. The black cells are cells with highly marked and rare rhyme structures both for nouns and verbs. The dark grey cells are cells with marked structures for both categories, without a strong N-V contrast, light grey cells are cells that show clear nominal-verbal contrasts, the white cells are cells that are common, unmarked syllabic structures both for verbs and nouns.

Table 5: Results Phonotactic Learner for monosyllabicity

rhyme		final consonantal cluster					
		zero	son	obstr	son + obstr	obstr + obstr	son+son
vowel of the syllable	lax monophthong	N	10,705	0	0	0	0
		V	11,971	0	0	0	1,101
		diff	-1,266	0	0	0	2,877
	tense monophthong	N	3,435	0	0,229	0	0,741
		V	2,857	0	0	5,409	4,912
		diff	0,578	0	0,229	-5,409	-4,171
	diphthong	N	1,568	0,484	0	1,652	2,337
		V	3,966	2,944	0	6,106	9,782
		diff	-2,398	-2,46	0	-4,454	-7,445
							-1,237

zero = no final consonant, son = sonorant, obstr = obstruent, N = noun, V = verb, diff = the weight for nouns minus the weight for verbs

The results show that there is a narrow field of prototypical, ideal syllables which are fully acceptable both as verbal and nominal stems (i.e. the white cells). These cells contain a constraint weight of zero value, which means that there is no constraint. Both as verbs and as nouns, these syllable structures are common in the Dutch lexicon. These syllables consist of a lax or tense vowel followed by a single consonant (as in *kat* 'cat' or *zoen* 'kiss'), a diphthong followed by an obstruent, as in *dijk* 'dike' or a lax vowel followed by a sonorant plus an obstruent (as in *werk* 'work'). To be very precise, for nouns with rhymes consisting of a tense monophthong plus an obstruent (as in *zoek* 'search') we did not measure a zero weight; we measured a constraint weight of 0,229. However, this value is still very low and can be interpreted as an unmarked, unrestricted syllabic configuration for Dutch. Given the very low value, it is also difficult to establish on the basis of these data whether the frequency of the lemmas involved has played a role. We propose to approach such lower numbers (i.e. values around 0,5 or lower) with some caution.

We measured, both for nouns and verbs, very high constraints for rhymes consisting of a lax monophthong not followed by a single consonant (as in the pseudo-word [da]) and for diphthongs followed by two sonorants (as in the pseudo-word *kourn*), represented by the black cells. We can conclude that these syllable types are illicit for Dutch, which is in fact commonplace knowledge.

There are a few cells which show no constraints at all for nouns (weight = 0), whereas they do for verbs (weight > 0). Lax monophthongs followed by two obstruents (as in *heks* 'witch') is an unmarked syllable type for nouns (weight = 0), but it is slightly marked for verbs (weight = 1,101). The same holds for a lax monophthong followed by two sonorants (as in *darm* 'bowel'), the weight for nouns is again 0, the weight for verbs is 2,877. Tense monophthongs followed by a sonorant plus an obstruent (as in *kaart* 'card') is unmarked for nouns (weight = 0), but considerably marked for verbs (weight = 5,409).

A few cells show a very mild constraint for nouns ( $1 > \text{weight} > 0$ ), but a remarkably stronger constraint for verbs. A tense monophthong plus two obstruents (as in *schaats* 'skate') is slightly marked for nouns (weight = 0,741), but considerably marked for verbs (weight = 4,912). For a rhyme consisting of a diphthong plus a sonorant (as in *pijn* 'pain'), we measured a very low weight for nouns (0,484), but a clear markedness weight for verbs (weight = 2,944). For this type of rhyme we would like to remark that the markedness value for the nouns is so low that the markedness should be considered with some caution. However, the contrast between the verbs and the nouns for this type is clear and undeniable (difference = 2,46). Diphthongs not followed by a single consonant (as in *dij* 'thigh') are also somewhat marked for nouns (weight = 1,568), but considerably more marked for verbs (weight = 3,966).

Then there are cells for which the syllable type is clearly marked for nouns, but the value for verbs tends to go towards nearly non-occurring. This holds true for tense monophthongs followed by two sonorants (as in *hoorn* 'horn'), with a weight of 2,977 for nouns and a weight of 6,385 for verbs. It is also what we observed for diphthongs followed by a sonorant plus an obstruent, as in *eind* 'end' (nominal weight = 1,652, verbal weight = 6,106) and for diphthongs followed by two obstruents, as in *beits* 'stain' (nominal weight = 2,337, verbal weight = 7,445).

Finally, a rhyme that consists of only a tense monophthong not followed by a consonant is marked for both nouns and verbs (dark grey cells). Surprisingly, the weight for nouns (3,435) is this

time higher than it is for verbs (2,857). The difference is rather small (0,578) and we are uncertain whether it is reliable.

In sum, for many cells we have measured a clear noun-verb contrast in the Dutch lexicon. There is a center field in the table which hosts both unmarked verbs and nouns. The edges of the table are clearly less marked as nouns. It is more common for a Dutch noun to have a richer syllable type than it is for a Dutch verb.

We would like to propose that these results are an effect of three syllabic phenomena that result in markedness. The idea is then that nouns are less affected by these markedness constraints than verbs. Firstly, extra-syllabicity is of course a marked situation. The Dutch rhyme, which contains at most three time units, can be followed by two more time units which occupy an extra-syllabic position.

Secondly, dispersion seems to be a factor. Dispersion refers contrasts in sonority. Consonantal clusters are better when the contrast in sonority between the consonants is clearer. Final combinations of obstruents and obstruents or combinations of sonorants plus sonorants are more marked than sonorant-obstruent sequences, as the contrast in sonority is lower. Perhaps the effect also applies to diphthongs plus sonorants, as one might consider the semi-vowel in the second time unit of the diphthong as a semi-consonantal sonorant.

Finally, Dutch syllables tend to have final consonants. Open syllables are not excluded, but they are certainly marked. For convenience's sake, we have indicated which constraint is relevant for which cell in table 6:

Table 6: markedness in the syllable

rhyme		final consonantal cluster					
		zero	son	obstr	son + obstr	obstr + obstr	son+son
vowel of the syllable	lax monoph.	open syllable				extra-syll., dispersion	dispersion
	tense monoph.	open syllable			extra-syll.	extra-syll.	extra-syll., dispersion
	diphthong	open syllable	disper-sion(?)		extra-syll.	extra-syll.	extra-syll., dispersion

monoph = monophthong, zero = no final consonant, son = sonorant, osbtr = obstruent, extra-syll. = filled extra-syllabic time units

It is clear that the cells which are not affected by these three markedness constraints are those cells which are unmarked as verbal stems in the Dutch lexicon. The Dutch lexicon thus shows that verbs tend to have a more ideal, prototypical syllable type for Dutch, whereas nouns resist markedness constraints better.

After measuring the restrictions for rhymes, we also measured onset restrictions. These are summarised in table 7. We used the same colouring as for Table 5: white cells indicate that there is

no difference between nouns and verbs, light grey cells indicate a contrast between nouns and verbs, and dark grey cells are cells with marked structures for both categories.

Table 7: Results Phonotactic Learner for onsets

onset			initial consonantal cluster					
			zero	son	obstr	obstr + son	obstr + obstr	son+son
vowel of the syllable	lax monophthong	N	2,255	○	○	○	○,109	8,035
		V	7,875	○	○	○	1,251	8,85
		diff	-5,62	○	○	○	-1,142	-0,815
	tense monophthong	N	2,055	○	○	○	1,156	8,335
		V	4,939	○	○	○	0,876	9,153
		diff	-2,884	○	○	○	0,28	-0,818
	diphthong	N	2,547	0,724	0,28	○	2,68	7,581
		V	7,729	1,108	0,791	0,028	1,549	8,393
		diff	-5,182	-0,384	-0,511	-0,028	1,131	-0,812

zero = no initial consonant, son = sonorant, obstr = obstruent, N = noun, V = verb, diff = the weight for nouns minus the weight for verbs

It is immediately clear that there are some noun-verb contrasts. Most notably, it is much more marked for a verb to have no onset than it is for nouns. For verbs a zero onset plus a lax monophthong (as in *as* ‘ashes’) the constraint weight (7,875) suggests that it is (nearly) non-occurring, whereas the markedness for nouns is much lower (weight = 2,255). We observe the same pattern for zero onsets plus diphthongs, as in *eik* ‘oak’, where the verbal constraint is 7,729 and the nominal constraint is 2,547. Zero onsets plus a tense monophthong, as in *aap* ‘monkey’ are clearly marked, but not excluded for verbs (weight is 4,939). The same syllable structure is clearly less marked for nouns (weight = 2,055).

For most other cells we either measure complete unmarkedness (i.e. zero constraints) for both verbs and nouns or very low differences (i.e. a difference around 0,5 or lower). We cannot confirm with confidence that these low values are reliable. For sequences of two sonorants we measure weights that indicate illicitness across the board (all weights for ‘sonorant plus sonorant’ are higher than 7,5).

For two obstruents followed by a lax vowel, as in *stap* ‘step’, the nouns are unmarked (weight = 0,109), the verbs are slightly marked (weight = 1,251). In contrast, for onsets which consists of two obstruents followed by a diphthong, as in *stijg* ‘raise’, we measure, surprisingly enough, less markedness for verbs (nominal weight = 2,68, verbal weight = 1,549). We do not yet have an explanation for this observation.

It is clear that the results for onsets are less outspoken than for rhymes, with one exception: both for nominal and verbal stems it is preferable to have a filled time unit in the onset, i.e. to have an initial consonant, but the requirement is of much more importance for verbal stems than it is for nominal stems.

This concludes our research of contrasts in the syllable structure of monosyllabic verbal and nominal stems in the Dutch lexicon. It has become clear that the Dutch lexicon certainly shows contrasts. Verbal stems are much more constrained by extra-syllabicity and dispersion and their syllables require clear consonantal contours, both in the onset and in the rhyme. In the next section we address the question whether the *number* of syllables differs for both categories as well.

#### 11.5.4.2 The number of syllables

First consider the manually counted distribution of nominal and verbal stems according to the number of syllables in our lists. We give the distribution first for all nominal and verbal stems. We then give the distribution for those stems that do not contain a schwa, to show the distribution of polysyllabic words with several full syllables.

Table 8: distribution of N and V according to number of syllables

	<b>total</b>	<b>1s</b>		<b>2s</b>		<b>3s</b>		<b>4s</b>		<b>5s</b>	
		<b>n</b>	<b>%</b>								
<b>all V</b>	<b>1401</b>	872	62,24	478	34,12	51	3,64	0	0	0	0
<b>all N</b>	<b>2939</b>	1207	41,07	1194	40,63	428	14,56	103	3,5	7	0,24
<b>V, no schwa</b>	<b>951</b>	872	91,69	79	8,31	0	0	0	0	0	0
<b>N, no schwa</b>	<b>2212</b>	1207	54,57	691	31,24	259	11,71	51	2,31	4	0,18

N = noun, V = verb, s = syllable(s), n = absolute number, no schwa = stem does not contain a schwa

It is immediately clear that the verbal lexicon and the nominal lexicon show a different distribution. The numbers differ greatly. Note that bisyllabic verbal stems with two full syllables are not excluded from the lexicon, but they are rare. More than two full syllables is excluded for verbal stems.

Next consider the results of the Phonotactic Learner to measure constraint weights for polysyllabicity. Now, one cannot simply formulate constraints on the number of syllables. Using the Phonotactic Learner, one is limited to constraints that measure sequences of phonemes (and word-boundaries). We therefore formulated constraints that measured the weights of vowels immediately following one another as in the word *theater* ‘theater’, of vowels following one another with a single intervening consonant, as in the word *tomaat* ‘tomato’ and with two intervening consonants, as in *garnaal* ‘shrimp’. The goal was not to measure a specific part of the syllable for these words. The explicit listing of syllabic material was just a work-around to make it technically possible to measure the constraint on the presence of more than one syllable. We measured both sequences of full vowels (described as V) and alternations between full vowels and schwa. The results are summarised in table 9. Unmarked constraints are white, (near-)absolute constraints are black and marked constraints are grey.

Table 9: constraints on polysyllabicity for nouns and verbs

Phonotactic sequence	N weight	V weight	Example
V-V	3,662	9,609	theater 'theater'
V-C-V	o	3,485	tomaat 'tomato'
V-C-C-V	o	3,269	garnaal 'shrimp'
V-schwa	4,782	8,268	sluier 'veil'
V-C-schwa	o	0,876	bode 'messenger'
V-C-C-schwa	o	o	lente 'spring'
schwa-V	8,952	7,45	beaam 'confirm'
schwa-C-V	o	o	gebed 'prayer'
schwa-C-C-V	o	o	verwen 'spoil'

V = vowel, C = consonant, N weight = weight for nouns, V weight = weight for verbs

The constraint weights for two full syllables with one or two consonants between the vowels (VCV, as in *tomaat* 'tomato' and VCCV as in *garnaal* 'shrimp') represents the most common structure for polysyllabic words. For nouns this structure is fully acceptable (weight = o), but for verbs we measure markedness (VCV weight = 3,485, VCCV weight = 3,269). It is a marked situation, but it is not excluded. We noticed this already in the distributional percentages (see table 8): 8% of the morphologically simplex Dutch verbal stems has two full syllables. It is clearly a marked situation, but it is certainly not excluded either.

Both for nouns and for verbs there is no constraint on combining a reduced syllable and a full syllable in either order, as long as one finds one or more consonants in between the syllables' vowels (V-C-schwa as in *bode* 'messenger', V-C-C-schwa as in *lente* 'spring', schwa-C-V as in *gebed* 'prayer', schwa-C-C-V as in *verwen* 'spoil') (all these weights are o, with the exception of the verbal weight for V-C-schwa, which is 0,876).

Two consecutive full vowels (VV) (as in *theater* 'theater') is marked for nouns (weight = 3,662), but illicit for verbs (weight = 9,609). A sequence of a full vowel plus a schwa (V-schwa, as in *sluier* 'veil') is highly marked for nouns (weight = 4,782), but (nearly?) illicit for verbs (weight = 8,268). Sequences of a schwa plus a full vowel, as in *beamen* 'to confirm', are very heavily marked for both categories and thus almost illicit (schwa-V: nominal weight = 8,952, verbal weight = 7,45).<sup>186</sup>

#### 11.5.4.2 The number of syllables in German

Before we proceed, I would like to discuss an important observation. We manually counted the distribution of nominal and verbal stems according to the number of syllables for German words as

<sup>186</sup> The fact that the verb has a slightly better score than the nouns for this type is to be ascribed to formally complex verbs that etymologically contain the prefix *be-*, but which are not perceived as morphologically complex today (such as *beamen* 'to confirm').

well, on the basis of the CELEX corpus. The CELEX corpus tags German lexical items, interestingly enough, for being the member of a conversion pair or not and they are tagged for the direction of the conversion. The tagging allowed me to compare constraints for verbs that are traditionally seen as denominal, such as *to shower*, and to count the number of syllables for such verbs. The results are given in the table below, together with the results for nouns and verbs in German in general. ‘No schwa’ means for German words that do not contain a schwa or a low schwa.

Table 10: distribution of N and V according to number of syllables

	<b>total</b>	1s	2s	3s	4s	5s	6s
	n	%	%	%	%	%	%
<b>all V</b>	1039	69	27	3	0,67	0,29	0
<b>all N</b>	3217	28	53	14	4	0,62	0,12
<b>V → N</b>	498	62	36	2	0	0	0
<b>V, no schwa</b>		88	10	2	0,37	0,37	0
<b>N, no schwa</b>	<b>2212</b>	50	34	12	4	0,52	0,12
<b>V → N, no schwa</b>		93	7	0,9	0	0	0

V = verb, N = noun, no schwa = the stem does not contain any reduced vowels, V → N = deverbal nominal conversion

Considering the number of syllables, it is clear that these verbs resemble the verbal pattern rather than the nominal pattern. The productive process thus takes into account the phonological constraints of the resulting category, rather than the original category. This holds true when compared both to all nouns and verbs and to those without schwas.

In a root-based framework this would mean that only roots that pattern with verbal syllabic structure are likely candidates to realise a root node in a verbal structure. Importantly, this shows that the interface relation that relates categories to syllable structure and vice versa is a relation between syntax and phonology rather than one between vocabulary and phonology. Roots seem to be subject to the constraints of the category in which they occur syntactically.

### 11.5.5 Conclusion

In this section, we have seen that the syllable structures of nominal and verbal stems differ in Dutch on the basis of computational corpus research. For verbs, the following constraints are much more severe than for nouns: consonantal contours, dispersion, no extra-syllabicity and there is a dispreference for more than one full syllable with an absolute constraint against more than two full syllables. We conclude that Trommelen’s observations hold true for the Dutch lexicon. In the next

section we argue that the native speaker's language competence has access to this N-V contrast. In other words, it is not only an observable fact, it is also psychologically real. Interestingly enough, however, we will establish that the marked cells for the native speaker's intuition do not coincide precisely with the cells that are marked in the Dutch lexicon.

## 11.6 Experiment: the psychological reality of the contrast

### 11.6.1 Introduction

Having established that the Dutch lexicon shows phonological noun-verb dissimilarities, we wanted to test whether this knowledge is part of the native speaker's phonology. The experiment by Don and Erkelens (2006) already indicated that this indeed seems to be the case. Yet, their results should be considered with caution, due to the methodological caveats of their study. We used the original design of this experiment, but altered it in two respects. Firstly, we increased the scale, both in the number of test items and in the number of participants. Secondly, we repaired some methodological caveats, which allowed for a more controlled experiment. More specifically, we controlled for the pseudo-words lexical neighborhood. This allowed us to control for word associations: we wanted to make sure participants based their judgements on the pseudo-words syllabic structure and not on its resemblance to an existing word. We also did not formulate the question to the participant at a linguistic meta-level. (Recall that they asked their participants: "Do you think this is a noun or a verb?").

The goal of the experiment was to test whether native speakers have judgments on the category of a pseudoword. We therefore presented them pseudowords and each item was given together with an indefinite determiner (a nominal context) and with a first person pronoun as a subject (a verbal context), as in the following example:

- (1) zeim
  - ik zeim
  - een zeim

Native speakers were asked to indicate whether they had a preference for *ik zeim* versus *een zeim*. We varied the syllable structure and the number of syllables and as such, we were able to test whether certain syllabic configurations were seen as more verbal or more nominal than others. Below, we report in more detail on the methodology and the results. It will become clear that, indeed, the native speaker's phonological competence is aware of noun-verb dissimilarities.

### 11.6.2 Methodology

#### 11.6.2.1 Materials

To generate test items, we wrote a script to automatically generate words with syllable structures that observed the principles of Dutch prosody. We thus generated a set of words that contained both existing words and non-existing forms that could have been Dutch words. We then used the Dutch Clearpond database, i.e. Dutchpond, to select from this set of words, the non-existing forms

with no orthographic neighbors<sup>187</sup>. In other words, we selected pseudowords that differed in at least two graphemes from existing words, to control for the lexical neighbourhood.. We created words of 33 different types of syllabic configurations and per type we selected 20 test items when possible. For some syllable types, we had fewer test items, because the syllable type is too densely populated with existing words. For the following monosyllabic configurations we thus worked with less test items: for syllables with a tense vowel + glide (as in *gooi* ‘throw’ and *leeuw* ‘lion’) we had 19 items, for syllables with a diphthong plus a sonorant (as in *brein* ‘brain’) we had 10 items, for diphthongs not followed by a consonant (as in *kou* ‘cold’) we had 10 items and for tense vowels not followed by a consonant (as in *zee* ‘sea’) we had also 10 items. For bisyllabic words of which the second syllable contained a lax or tense vowel followed by two consonants, we did not design separate syllable types for the sequences sonorant + sonorant, sonorant + obstruent and obstruent + obstruent. Instead, we grouped these three cells as a single syllable type. The motivation was that these cells are only marginally attested in Dutch and we decided not to burden the size of the experiment by keeping them separated. In total, we had 649 test items. Each test item consisted of 4 to 8 graphemes. We did not test for onsets, but nearly all test items had filled onsets.

As mentioned above, we presented native speakers pseudo-words and each item was given together with an indefinite determiner (a nominal context) and with a first person pronoun as a subject (a verbal context). The order between the nominal context and the verbal context was automatically given in a random order by the questionnaire software. The participants were asked to select the option that felt most natural to them. They were told to pick just anything when both answers were equal options.

The experiment was designed as a large-scale online experiment. The test items were equally distributed over 10 lists, with pseudo-random order. For each condition, each participant saw 10% of the available test items. For example, if we had 20 test items for a given syllable type, they would be distributed over the ten lists, such that each list had two test items of that condition. All participants thus were presented an equal part of each condition. We had in total 649 test items, so nine lists contained 65 test items and one list contained 64 test items. Each list became a questionnaire in SurveyMonkey. The lists were hidden behind a link that connected to a piece of code that would send people randomly to any of these lists. Responses to the separate lists were thus gathered simultaneously. We closed a list when we had 60 responses of people who qualified for our selection criteria.

#### *11.6.2.2 Procedure*

Participants were first trained doing the task with nine real words as examples and they were given feedback. Four of the practice words were unambiguously nominal (*heks* ‘witch’, *reeks* ‘series’, *kussen* ‘pillow’<sup>188</sup>, *kreeft* ‘lobster’), four were unambiguously verbal forms (*sticht*<sup>189</sup> ‘establish’, *fluister* ‘whisper’, *meng* ‘mix’, *oefen* ‘practice’). One of these words was ambiguous (*duw* ‘push’) and it was

---

<sup>187</sup> It is impossible to count the number of *phonological* neighbours of non-existing words, as, of course, no phonological transcriptions of non-existing words are available in the database. We therefore worked with *orthographic* neighbours.

<sup>188</sup> *Kussen* could be an infinitive (‘to kiss’), but is unacceptable in the 1sg form. This word was intentionally in the exercise so that participants should only select *ik+word* if the word was acceptable to them as an ‘*ik*-form’.

<sup>189</sup> Some participants reported in the comment section they knew *sticht* as a noun as well. According to the WNT it is used as a noun to refer to a road around Oudewater and Montfoort and as an archaism to refer to a monastery in southern dialects.

explained in the feedback that for such a form either answer would be fine. Results of participants who made errors on this initial exercise were not used.

After having practised with real words, participants received five practice questions with pseudowords. Participants were explicitly told they were still practising. After this exercise, it was announced that the actual experiment would start and 65 test items followed. Afterwards they also received some personal questions, which we used to exclude non-native speakers of Dutch, linguists, people who grew up bilingually, people who lived in a region where no Dutch is spoken for more than 2 years and children.

#### *11.6.2.3 Participants and collected data*

A link to the questionnaires was distributed via *Taalpost*, a popular-scientific newsletter read by people who like to be kept up-to-date on news relating to the Dutch language. It is sent out twice per week and at the time of this project it was read by more or less 80 000 people. The results were gathered and stored anonymously. We gathered data from 600 participants, i.e. 60 participants per list.

For syllable types for which we had 20 test items, we gathered 1200 judgments in total from 600 participants (10 lists x 60 participants per list, 2 test items per syllable type per participant), for the three syllable types for which we only had 10 test items, we had 600 judgments in total from 600 participants (10 lists x 60 participants per list, 1 test item per syllable type per participant).

### **11.6.3 Results**

Given what we know about the Dutch lexicon thanks to the Phonotactic Learner, we expected an equal distribution between nounhood and verbhood for the central cells, with a bias towards verbhood. Note that less test items were expected to qualify as good verbs and participants tend to spread their answers equally over the options given. As such, cells of which the syllables could be recognised as verbal were expected to have a somewhat skewed distribution, i.e. we were expecting something like a 60/40 distribution in the favour of verbs. For the peripheral cells, we were expecting a higher number for nounhood. We did not expect absolute percentages (i.e. a score close to a 0/100 distribution) for two reasons. Firstly, we have seen that the verb-noun dissimilarities are better to be understood in terms of strong probabilities rather than as absolute constraints. Secondly, we have noted in former online questionnaires that these seldomly result in absolute results.

Table 11 below summarises the percentages we gathered for monosyllabic forms. Black cells were not included in the questionnaire as these correspond to illicit syllable types. Grey cells are cells that were judged dominantly nominal, the darker the cell, the stronger this effect (light grey: % N between 50 and 61, middle grey: % N between 62 and 70, dark grey: % N higher than 70). White cells were judged both as verbal and nominal (% N < 50). For each syllable type, we present the nominal values and the verbal values as percentages and we give the standard deviation.

Table 11: Results judgement task: monosyllabic words

Monosyllabic forms			final consonantal cluster of the syllable						
vowel of the syllable			zero	son	obstr	son+ obstr	obstr + obstr	son+ son	C+C+s/t
	lax monoph	% N		37,8	36,75	40	59,1	45,3	73
		% V		62,3	63,25	60	40,9	54,7	27
	tense monoph	SD		14,3	10,7	12,9	11,6	11,2	9,2
		% N	43,5	38,17	38	69,8	64,2	56,4	73,2
		% V	56,5	61,8	61,9	30,2	35,8	43,6	26,8
	diphthong	SD	7,35	10,4	11,9	14	14,5	17,3	8,8
		% N	34,3	41,2	41	60,8	67,2		
		% V	65,7	58,8	59	39,2	32,8		

monoph = monophthong, zero = no initial consonant, son = sonorant, obstr = obstruent, C+C+s/t = sequence of two consonants plus /s/ or /t/, % N = percentage of the cases the nominal structure was chosen, % V = percentage of the cases the verbal structure was chosen, SD = standard deviation amongst percentages of the test items for the given syllable type

All white cells are fully accepted as verbs. They include all syllable types with open syllables and all syllable types with one final consonant (% V > 50). Syllables with a lax monophthong followed by a cluster of two consonants are seen as possible verbs if the first consonant of those two consonants is a sonorant, as in *werk* ‘work’ and *worm* ‘wurm’. None of these syllable types contain any extra-syllabic positions. Possible verbhood decreases if a lax vowel is followed by two obstruents (as in *heks* ‘witch’, % V = 40,9), even though this syllable types still does not contain an extra-syllabic position.

The other grey cells contain extra-syllabic positions and it is clear that possible verbhood decreases according to the native speaker. Diphthongs followed by a sonorant plus an obstruent, as in *deins* ‘recoil’, (% V = 39,2) and tense vowels plus two sonorants, as in *hoorn*<sup>190</sup> ‘horn’ (% V = 43,6) are mildly less preferred as verbs. Verbhood goes further down if the tense vowel is not followed by two sonorants, but by a sonorant plus an obstruent, as in *moord* ‘murder’ (% V = 30,2), or two obstruents, as in *feest* ‘party’ (% V = 35,8). Diphthong plus two obstruents, as in *beits* ‘stain’, are in the same league for possible verbhood with %V = 32,8. Possible verbhood drops below 30% as soon as a form contains a final consonantal cluster of three consonants, as in *worst* ‘sausage’ and *koorts* ‘fever’.

Next, consider the results for the bisyllabic words (Table 12):

<sup>190</sup> A word such as *hoorn* ‘horn’ has an extra-syllabic position which is realized by a consonant other than /s/ or /t/.

Table 12: Results judgement task: bisyllabic words

Bisyllables			final consonantal cluster						
vowel of the syllable	lax monoph	% N	zero	son	obstr	son+ obstr	obstr + obstr	son+son	C+C+s/t
		% V		16,4	20,3	14,7			
		SD		12,8	7,5	9,6			
	tense monoph	% N	84,75	80	82	85,75			
		% V	15,25	19,9	17,9	14,25			
		SD	13,4	10,8	10,5	9,2			
	diphthong	% N	79,1	82	81				
		% V	20,9	17,9	19				
		SD	8,5	10,4	13,5				
	schwa	% N	83,5 [əm] [əl]/[ər] ]	78,3 60,3	83,25	86,3			
		% V	16,5 [əm] [əl]/[ər] ]	21,7 39,7	16,75	13,7			
		SD	7,6 [əm] [əl]/[ər] ]	11 17	5,8	6,8			

monoph = monophthong, zero = no initial consonant, son = sonorant, obstr = obstruent, C+C+s/t = sequence of two consonants plus /s/ or /t/, % N = percentage of the cases the nominal structure was chosen, % V = percentage of the cases the verbal structure was chosen, SD = standard deviation amongst percentages of the test items for the given syllable type

Only, if the second syllable of a bisyllabic form is [əl] or [ər] possible verbhood is around 40%. Otherwise, for bisyllabic forms, possible verbhood is consistently around 20% or lower. If the final syllable ends in more than one consonant, it even consistently drops below 15%. Recall that one rarely measures absolute judgements in large-scale online judgement tasks. We would expect that in a more controlled laboratory setting, one would measure an even more outspoken preference for nounhood for such cells.

Statistical tests confirm the significance of the number of syllables and extra-syllabic positions.

Bisyllabic pseudo-words ( $M = 19.128$ ,  $SE = 0.700$ ) are judged far less often as verbs than monosyllabic words ( $M = 48.639$ ,  $SE = 0.986$ ) by native speakers. This difference,  $29.511$ , BCa 99% Confidence Interval [26.508, 32.947] is significant  $t(607.592) = 24.395$ ,  $p < 0.001$ . (Equal variances not assumed, Levene test  $p < 0.001$ ). Furthermore, within the group of monosyllabic types, syllables with extra-syllabic positions ( $M = 33.607$ ,  $SE = 1.230$ ) are judged far less often as verbs than syllables without extra-syllabic positions ( $M = 58.708$ ,  $SE = 0.09$ ) by native speakers. This difference,  $25.101$ , BCa 99% Confidence Interval [-29.011, -21.375] is significant  $t(347) = -16.748$ ,  $p < 0.001$ . (Equal variances assumed, Levene test  $p = 0.258$ ).

#### 11.6.4 Discussion

From the results it is immediately clear that noun-verb dissimilarities are represented in the native speaker's phonology. Certainly not all cells are judged as equally verbal-nominal for monosyllabic forms. For bisyllabic forms native speakers have a very strong preference for nominality.

It would be a reasonable assumption that speakers are aware of statistical patterns of their language and that this is the effect we have measured. Interestingly, however, the judgments of the native speakers are not exact reflections of statistical knowledge of their language. This becomes clear when we compare the results of the Phonotactic Learner and the experiment results. For convenience's sake, we have summarised the results in the following tables (tables 13 and 14) for the monosyllabic forms. The black cells are non-attested syllable structures, grey cells are predominantly nominal cells, white cells are both verbal and nominal.

Table 13: Results Phonotactic Learner

monosyllables		final consonantal cluster					
		zero	son	obstr	son + obstr	obstr + obstr	son+son
vowel of the syllable	lax monoph	black					
	tense monoph						
	diphthong						black

Table 14: Results judgement task

Monosyllables		final consonantal cluster					
		zero	son	obstr	son+obstr	obstr + obstr	son+son
vowel of the syllable	lax monoph	black					
	tense monoph						
	diphthong						black

monoph = monophthong, zero = no initial consonant, son = sonorant, obstr = obstruent

It is clear that what native speakers experience as a possible verbal structure shows overlap, but does not coincide with what is statistically a prototypical verbal structure in the lexicon. More specifically, monosyllabic forms with a diphthong not followed by a consonant, as in *kei* 'boulder', are strongly perceived as possible verbal forms by native speakers (diphthong plus zero C = 65,7% verbhood), even though the lexicon does not provide the statistical grounds to base this decision on at all. The constraint weights for nouns of this format was 1,568 and for verbs 3,966. If speakers based their judgment solely on the lexicon, they should have had an outspoken preference for nounhood. A diphthong plus a sonorant, as in *duim* 'thumb' or *pijn* 'pain', is judged as verbal in 58,8% of the cases by native speakers. Yet, the constraint weight for such verbs was 2,944 versus 0,484 for the nouns. A monosyllabic form consisting of a lax monophthong plus two sonorants, as in *worm* 'wurm', is judged as verbal in 54,7% of the cases by native speakers. This judgement again does not reflect subconscious statistical knowledge of the Dutch lexicon, as nouns are fully unmarked for this type (weight = 0), whereas verbs are clearly marked (weight = 2,877). These results are exciting and interesting. The intuition of the native speakers is not solely steered by their lexical knowledge. Something else is at work here that we at this point fail to understand, but which we intend to research further in the future.

Other results do correspond between the Phonotactic Learner and the judgement task. The monosyllabic forms that appeared to be unmarked as verbs according to the Phonotactic Learner (i.e. the white cells in table 14) are indeed seen as possibly verbal by native speakers. Monosyllabic forms that have a tense monophthong not followed by a consonant were both marked as verbs (weight = 2,857) and as nouns (weight = 3,435), with the verbs being slightly better than the nouns. As expected, people do tend to judge these forms as possibly verbal (56,5% verbhood).

We saw that 8% of the morphologically simplex Dutch verbs are bisyllabic and we also saw that polysyllabicity is marked according to the Phonotactic Learner for verbs (see table 9, weight of VCV verbs = 3,485; weight of VCCV verbs = 3,269), whereas it is unmarked for nouns (both for VCV and VCCV nouns the weight is 0). Bisyllabicity is clearly the exceptional situation for verbs, but it is not excluded. Native speakers have a strong preference for nounhood for bisyllabic forms, the percentages range between 78% and 86% (with the exception of forms that end in schwa plus l/r).

These are important results. They indicate that there is an inherent relation between syllable structure and a word's categorial status for the native speaker of Dutch in a way that transcends mere statistical phonotactic awareness. There must be an interface relation for Dutch we do not fully understand at this point and which we aim to develop further on in the future.

## 11.7 Optimal paradigms?

In section 4.4 we discussed the hypothesis from Optimal Paradigms that all phonological noun-verb dissimilarities can be ascribed to inflectional differences. McCarthy (2005) also proposed that candidates for evaluation are entire paradigm sets rather than individual items. This section aims to test both hypotheses for Dutch. We will first investigate the plausibility of ascribing the noun-verb contrast to inflection in Dutch. We will subsequently present an experiment that tested whether people indeed evaluate paradigm sets.

### 11.7.1 Nominal versus verbal inflection

Both nominal and verbal inflection have suffixes that attach extra-syllabically to the word's final syllable and suffixes that simply form syllables on their own. These suffixes are schwa syllables:

(2) twee bel-en	(3) we bel-en
two bell-PLURAL	we call-INFLECTION
'two bells'	'we call'

Suffixes that form a schwa syllable on their own are of less interest to the present discussion. We have seen in section 5.3 that both verbs and nouns tolerate an additional reduced syllable. Let us therefore focus on those suffixes that attach extra-syllabically to the nouns' or verbs' final syllables. The verbal domain has one such suffix: the ending of the second and third person singular of the present tense is a -t, as in *hij werk-t* 'he works'. The nominal domain has a plural ending that is an -s, as in *otter-s* 'otters'. There is, however, a crucial difference between the verbal and the nominal ending: the verbal -t is unavoidable if the native speaker wants to use a verb as a full-fledged finite verb. The nominal -s, on the other hand, competes with the plural marker -en. As such, it is avoidable when phonologically less optimal, as an alternative is available.

According to this logic, it follows that the maximum number of filled moraic positions for verbs could be one unit smaller than the number for nouns. After all, the verbs need to be able to 'grow' one time unit in case it will be used in the second or third person singular in the present tense:

(4)	noun	xxxxx
	verb	xxxx(t)

As a hypothesis, the idea that syllabic verb-noun contrasts stem from inflectional differences is not absurd for Dutch. The experiment shows that for monosyllabic words, the more the syllable grows qua time units, the more it is judged to be nominal. Extra-syllabic endings for bisyllabic words are also rather marked for Dutch, so as soon as a bisyllabic verbal stem would contain at least one consonant, an extra suffix -t would result in a highly marked form. The only observation that cannot be immediately captured is that also bisyllabic forms ending in a vowel tend to be seen as nominal. Yet, one could make a case that the extra -t would make the final syllable superheavy (*ik neurie* 'I hum' versus *jij neuriet* 'you hum'). Given that superheavy syllables tend to attract word stress, this might be a less optimal situation for bisyllabic verbs that have their word stress on the first syllable.

The concern that a verb can only be a useful verb when it inflects well could affect a language in two ways. Firstly, if the Optimal Paradigm hypothesis is correct that language speakers judge entire paradigm sets, it could make a stem overall less optimal. We discuss this possibility in the next section. Secondly, it could work as a silent force through the ages. It is plausible that verbs that are not great phonologically in the second and third person singular are less frequently used and may become less central members of the vocabulary or even disappear. If this were true, we should see this effect in the lexicon. Consider the results from the Phonotactic Learner (table 15). If the verb inflection would play a role, we would expect verb stems to behave as nouns that have one obstruent more, since they have to safe a place for the -t. We have indicated with arrows which verbal stems should correspond qua value to which nominal stems: the values of the verbs should correspond to those of nouns that are one obstruent richer, when we do not take into account verbal stems, but

verbs plus the ending -t (i.e. verbs plus one obstruent). The correspondence is actually not so great. Even though the reasoning makes sense, the numbers do not support it. Sure there are some unmarked cells which correspond to unmarked cells (i.e. a zero corresponds to a zero). What is more telling, is that the non-zero values do not correspond to a similar value.

Table 15: Phonotactic Learner: verbs + t?

rhyme		zero	obstr	obstr + obstr	son	son + obstr
lax monoph	N		o	o	o	o
	V		o	1,101	o	o
tense monoph	N	3,435	o,229	o,741	o	o
	V	2,857	o	4,912	o	5,409
diphthong	N	1,568	o	2,337	o,484	1,652
	V	3,966	o	9,782	2,944	6,106

In short, the difference between verbal and nominal inflection might in principle result in more stringent requirements for the verbal stem. However, looking at the numbers, we cannot confirm that this effect shaped the Dutch lexicon. In the next section we present an experiment that tests whether people judge paradigm sets rather than single candidates.

### 11.7.2 The experiment design

If it were true that people judge entire paradigms rather than single candidates, forms that are part of more optimal paradigm sets should be judged better than forms that are part of less optimal paradigm sets. Within OP, this reasoning is usually put to use to derive paradigm uniformity: all members of the paradigm stand in a relation to one another and are all evaluated together. As such, a high-ranking faithfulness constraint that expresses a preference for paradigm uniformity can kick in easily. However, when paradigm uniformity is not at issue as is the case in the present discussion, the evaluation procedure developed by OP should still apply to other aspects of the forms to be evaluated. If OP is right, we should see paradigm effects when judging a single form.

Now consider that Dutch has degemination. As a result, when a Dutch verbal stem ends in a /t/ and an extra inflectional ending /t/ will be added at the underlying representation, the form will surface with a single /t/: *hij zit-t 'he sit-s'* becomes *hij zit 'he sits'*. This means that for verbal stems ending in a -t the syllable structure will not gain an extra time unit in the second and third person singular present tense. Ending in a -t is thus an advantage for more complex verbal stems: the forms will not become even more complex throughout the paradigm. Overall, paradigm sets for verbal stems ending in a -t will be thus be better than paradigm sets for other verbal stems in Dutch.

If it were true that native speakers judge entire paradigm sets rather than individual candidates, pseudo-verbs ending in a -t should be judged as being better than other verbs. This should be true even if we do not present the form in the second and third person singular. We presented native

speakers with both types of pseudo-words, i.e. words ending in a -t and words that end in a different consonant. Again, we asked people to judge whether they felt the pseudo-word was nominal or verbal by letting them pick between a nominal and a verbal syntactic context, as in the previous experiment:

(5) morfts

- ik morfts
- een morfts

Note that the verbal context contains again the first person singular. We predicted, of course, on the basis of the previous experiment that the more complex the syllable would become, the more the nominal context would be selected. However, if it were true that people judge entire paradigm sets, we expect this effect to be milder for verbal stems ending in a t: their paradigms do not grow by one position for the second and third person singular in the present tense and, as such, these forms should overall be better as verbs.

If people simply judge single candidates, this effect should be non-existent. It should not matter whether a pseudo-word ends in a -t when judged as a verb.<sup>191</sup>

### 11.7.3 Methodology

We tested 6 syllable types, as given in table 15 below. The white cells are the cells that we tested. For each cell in the table, we give some real-word examples for your convenience. The black cells are cells that are non-existent in Dutch. The grey cells were untestable cells. For the cell that contains a tense monophthong followed by zero consonants, we obviously cannot test the effect of a -t. For lax monophthongs followed by a single consonant there were no test items available without any orthographic neighbours ending in a /t/ or /k/, forms ending in a /p/ or /s/ both only had four test items available. We decided not to include this cell in the experiment. We worked with the following number of test items:

Table 16: test items

vowel	number of consonants	following consonants	real-word example	#test items
lax	2	sonorant + [p] or [k]	harp, hark	20
lax	2	sonorant + [s]	hars	7
lax	2	sonorant + [t]	hart	20
lax	3	sonorant + obstruent + [s]	wulps	20
lax	3	sonorant + obstruent + [t]	markt	20
lax	4	sonorant + obstruent + [t] + [s]	/	20

<sup>191</sup> We have checked using the Phonological Learner whether there is a constraint on Dutch verbal stems to end in a -t and there is no such constraint. Indeed, frequent every-day verbal stems may end in a -t in Dutch (*zit* 'sit', *wacht* 'wait', *laat* 'let'), this is not a marked situation.

Table 16: test items

vowel	number of consonants	following consonants	real-word example	#test items
<b>lax</b>	4	sonorant + obstruent + [s] + [t]	herfst	20
<b>tense</b>	1	[p] or [k]	piek, piep	20
<b>tense</b>	1	[s]	kies	10
<b>tense</b>	1	[t]	piet	20
<b>tense</b>	2	sonorant + [s]	kaars	20
<b>tense</b>	2	sonorant + [t]	kaart	20
<b>tense</b>	3	C + [t] + [s]	koorts	20
<b>tense</b>	3	C + [s] + [t]	oogst	20

For the pseudo-words with the tense vowel plus a single consonant (*piet*) it was *a priori* unclear to us whether pseudo-words ending in a [p], [k] or [s] would be the best comparisons for pseudo-words ending in a [t]. We therefore decided to include 10 test items ending in [p], 10 ending in [k] and 10 ending in [s]. The same reasoning is valid for syllables of the type *hars/harp/hark/hart* with the nuance that for the type ‘hars’ no more than 7 test items were available. For all other syllable types the /t/ already occupies an extra-syllabic position and it was clear that we had to compare the /t/ with the /s/: the /t/ and the /s/ are the only consonants that are acceptable in these positions.

We had 257 test items in total. The test items were equally distributed over 4 lists, with pseudo-random order. For each condition, each participant saw 25% of the available test items. For example, if we had 20 test items for a given syllable type, they would be distributed over the four lists, such that each list had five test items of that condition. All participants thus were presented an equal part of each condition. We had in total 257 test items, so three lists contained 64 test items and one list contained 65 test items. We copied the practice part of the previous experiment and the additional questions (see section 5.3). Again, the questionnaire was set up in Surveymonkey and distributed via Taalpost (see section 5.3).

We closed each list when 200 people had filled out the questionnaire. We excluded participants again according to pre-defined criteria: we excluded non-native speakers of Dutch, linguists, people who grew up bilingually, people who lived in a region where no Dutch is spoken for more than 2 years, children and people older than 80 years. Of the remaining participants, we kept the first 100 participants per list. We counted how often participants had judged a syllable type as verbal and we averaged the results for test items of the same syllable type.

#### 11.7.4 Results

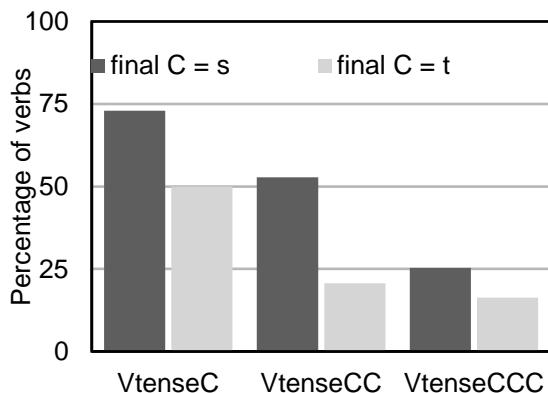
The following tables summarise the results. A comparison between forms ending in /s/ and those ending in an /t/ is always given in the graphs (graphs 1 and 2).

Table 17: Lax vowels  
(percentage of verb preference)

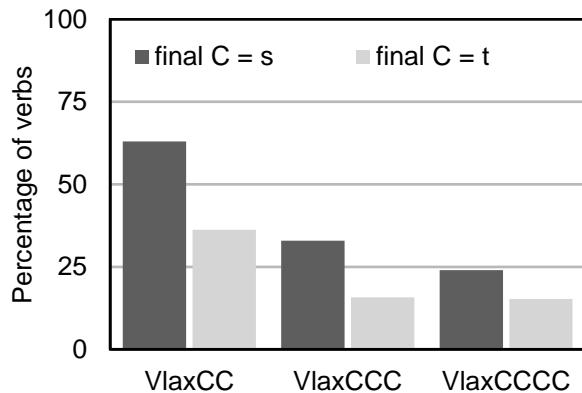
		number of final consonants		
		2	3	4
nature of final consonant	/p/ or /k/	67		
	/s/	63	33	24
	/t/	36	16	15

Table 18: Tense vowels  
(percentage of verbs)

		number of final consonants		
		1	2	3
nature of final consonant	/p/ or /k/	70		
	/s/	73	53	25
	/t/	50	21	16



Graph 1: Stems with tense vowels ending in s/t



Graph 2: Stems with lax vowels ending in s/t

We tested structures with tense monophthongs and lax monophthong, including syllables that contained a rhyme with three filled time units (VtenseC and VlaxCC). Three time units is the maximum number of time units the Dutch rhyme can contain, so any additional inflectional -t would already occupy the more marked extra-syllabic position. We also tested structures which already contained three time units in the rhyme plus a filled extra-syllabic position (VtenseCC and VlaxCCC). An extra -t would then deplete all available extra-syllabic positions. We also tested complete rhymes plus two filled extra-syllabic positions (VtenseCCC and VlaxCCCC). Such structures in fact have no position to host an additional -t. Recall that we compared forms that did not end in a -t and thus need to host an additional -t in the second and third person singular present tense, and forms of which the stems already end in -t and which thus do not will get an extra consonant when inflected.

Two patterns are clear. Firstly, as expected from the results of the previous experiment, the more complex the syllable becomes, the more forms are felt to be nominal, regardless of the final consonant. Secondly, forms ending in a consonant other than /t/ are seen as better verbs in the first person singular than those forms that do end in a /t/. Recall that the hypothesis was that if the Optimal Paradigms were right, the forms with the /t/ should be the better forms. This is clearly not borne out. Recall also that the Phonotactic Learner predicted that it should not matter for a verb whether it ends in a /t/ or not, as it is unmarked for a Dutch verb to end in a /t/. This expectation is also not borne out: the presence of a final /t/ clearly makes a pseudo-word more marked as a verb in the first person singular.

One may intuitively guess what happened: participants probably have decomposed the forms ending in a /t/ into stems plus -t. They may have recognised the -t as an inflectional and they therefore may have judged the forms as less suitable in the first person singular. They clearly judged this form as an individual form and deemed it less suitable for a first person singular. There is thus no evidence at all that they judged entire paradigm sets.

The aim of this entire section was to investigate whether the difference between verbal and nominal inflection can be the ground for the syllabic differences between nouns and verbs. We described that, in principle, this effect could be reached in two ways. It could be a force that shapes the lexicon through the ages. In section 7.1 we described that the inventory of the Dutch lexicon does not support this idea numerically. We then tested the hypothesis that is a force that operates online, during the processing of language as proposed by the Optimal Paradigms approach. The results of this experiment provide no evidence that people indeed evaluate entire paradigms.

We conclude that, even though looking at inflectional differences in principle was a promising track to account for the phonological noun-verb dissimilarities, we did not find any empirical evidence to support such a hypothesis.

## 11.8 Discussion

Before concluding this article, we would like to discuss a few considerations, including the probability-based nature of this work, the role of loanwords in the lexicon and the role of formal complexity of words.

Throughout the article it was clear that we are dealing with an empirical domain that adheres to tendencies and probabilities. Absolute generalisations do not apply. This fact should be captured by any future explanatory account. We would like to warn against the hasty conclusion that it is simply a given that the Dutch lexicon shows certain tendencies which can be explained by pointing at historical and etymological happenstance. We have seen that, indeed, the Dutch lexicon shows noun-verb dissimilarities which were picked up by the Phonotactic Learner. We have also seen that native speakers have intuitions on which syllable structure is to be associated with nouns or verbs. Crucially, however, the results of the Phonotactic Learner and the experiment did not coincide. The idea that the lexicon evolved in a certain way and that these statistical patterns are subconsciously picked up by the native speaker is too simple. The phonology, the lexicon and the syntax of the native speakers are interfacing in a way that allows the native speaker to connect certain syllable structure to word classes in ways that transcend statistical patterns in the lexicon. This result, together with the probability-based nature of the facts is the explanandum for future research.

Trommelen (1989) already observed that polysyllabic structures of which the final syllable does not contain a schwa are always originally loanwords. In fact, the list of polysyllabic nouns contains

many words that originally were loanwords. One could assume that what we described in the present article is simply loan phonology. Again, we would like to warn against hasty conclusions for four reasons. The first reason was just mentioned above: the experiment shows that native speakers have judgments that transcend the make-up of the lexicon. Simply pointing at the nature of the Dutch lexicon cannot capture this fact. Secondly, it only postpones the puzzle by one step as one may question why polysyllabic loanwords are fully accepted by the language in the nominal domain and not in the verbal domain. Thirdly, mostly these words may originally have been loanwords, but they are fully integrated morphophonologically. The examples do not show loan phonology (beyond their polysyllabicity) and they may select the plural marker *-en*, even though we know that recent loanwords would select the default plural marker *-s*. As such, there is no reasons to assume these words are considered as loanwords by the language user. Their foreign etymology is an obvious fact to the linguist, but it is not clear if it plays any role in the language system of the native speaker. Fourthly, Dutch is not the only language in which noun-verb asymmetries can be explained etymologically. The same is true for Hebrew (Bat-el 1994). This has not prevented Hebrew from playing a central role in the theoretical discussion on noun-verb asymmetries. In fact, one could ask precisely the same question for Hebrew: why are loanwords welcomed so much more easily in the nominal domain than in the verbal domain. In terms more suitable for the Semitic language, this question can be rephrased as: Why are verbs templatic, but nouns not necessarily so? Bat-el (1994) and Arad (2005) raise exactly this question.

For obvious reasons, Trommelen (1989) excluded morphologically complex data from her research and so did we. Of course, all compounds would by definition be polysyllabic. The same holds for most derivations, as most derivational affixes are syllabic in Dutch. Including morphologically complex words would only have distracted from the core observations. Trommelen (1989) tacitly did not include words that are only formally complex either. Formally complex words are words in which an affix-like syllable and a stem may be recognised on the basis of morphophonological grounds, even though no semantics can be attached to these separate elements. We decided to include formally complex verbs and nouns. Certainly the formally complex verbs form a significant part of the lexicon. We felt that excluding them would manipulate the verbal lexicon on becoming more monosyllabic and uniform than it is in reality. Yet, the syllabic structure of formally complex verbs (and nouns) is probably picked up by the native speaker and may signal syntactic information.

In sum, we believe that the fact that nouns and verbs in the lexicon differ and that the nominal domain accepts loanwords much more willingly should be seen as the consequence of a crucial difference between nouns and verbs at the interfaces rather than as the cause of this difference. Formal complexity, we believe, is an interfering issue which we did not address in this article, but which arguably plays a role in this domain.

## 11.9 Conclusion

This article aimed to flesh out Trommelen's (1989) observation that Dutch shows phonological noun-verb dissimilarities. She observed that verbal stems are most often monosyllabic and those single syllables also tend to be less complex than those of nouns. Don and Erkelens' (2006) experiment suggested that this knowledge is probably part of the native speaker's phonology.

We found that, indeed, the verbal part of the lexicon differs from the nominal part qua syllable structure. We were able to confirm that verbal stems contain less syllables and less complex

syllables than nominal stems through computational research. We fed lists of verbal stems and of nominal stems from the SUBTLEX-NL corpus to the Phonotactic Learner and we were clearly able to identify prototypical verbal syllable structures.

A short investigation into the syllable structure of German deverbal nominal conversion, which was possible thanks to the tagging in the German CELEX corpus, shows that such nouns pattern more closely with verbs than with nouns qua syllable structure. It seems that lexical items need to fit the syntactic structure in which they occur: syntax interfaces with phonology.

In a large-scale online experiment we were also able to find clear suprasegmental noun-verb oppositions. Native speakers do have an intuition on the word class of a pseudo-word on the basis of its syllable structure. Again, the intuitions are steered by the syllables' complexity and the number of syllables. However, the experimental results do not fully coincide with the corpus research. This indicates that something steers this process in the linguistic competence that transcends mere awareness of statistical patterns in the lexicon.

These results are fully in line with typological observations on noun-verb dissimilarities. As in other languages, the differences are situated at the suprasegmental level of the phonology and nouns are the more liberal category.

We investigated whether the cause of the noun-verb contrast could be ascribed to inflectional differences. We did not find any evidence that inflection had caused a cross-categorial lexical contrast and we did not find any evidence for the hypothesis that the paradigm as a whole can put pressure on the form of the stem either. We aim to investigate other hypotheses to account for the phenomenon in the future.

## 12 Conclusion

### 12.1 Introduction

The present bundle of chapters discussed Dutch morphosyntax and its interfaces within the framework of Distributed Morphology. It started with the question which modules and interfaces are involved for which data and how the modules interface. As to the question which module is responsible for which data, I have proposed systematically to assign as much structure as possible to the ‘rule-based’<sup>192</sup> modules, viz. syntax and phonology, rather than to ‘storage-based’ cognitive capacities, such as vocabulary. For example, I have decomposed the non-heads of Dutch primary compounds into roots and functional heads -if any- (chapters 2-4) and, as such, I have identified three types of primary compounding in Dutch on the basis of their syntax. Other examples are the syntactic decomposition of posture verbs in agentive and stative verbs and the maximal assignment of vowel length alternating allomorphy to phonology. For all these domains I have thus assumed structure rather than storage. I have also disentangled complex domains in which syntax, vocabulary and phonology interact in the chapters on epenthetic consonants and the relation between syllable structure and categories.

As to the question which interfaces interact with one another, I have contributed empirical observations relevant to the relation between morphology and syntax and the interfaces between morphosyntax and vocabulary, morphosyntax and semantics and morphosyntax and phonology, as discussed in more detail below.

### 12.2 Morphology and Syntax

The present bundle fully adopts and supports the view that morphology is syntax. The motivation is principled: I do not adopt structure-building modules beyond necessity. The chapters on compounding explicitly contribute to the view that words are built in syntax. As such, compounds, and words in general, can contain both syntactic heads (chapter 3, 4 and 10) and phrases (chapter 5).

Word-hood<sup>193</sup> is derived via incorporation (i.e. head movement). Despite the claim that morphology is syntax, word-hood is taken seriously. It is emphasized that the word has a different psychological status than the sentence: speakers feel that a word can ‘exist’, sentences, in contrast, are simply ‘produced’. Word-hood also seemed to be of importance as the relevant domain for allomorphy and it has been argued empirically in chapter 9 that incorporation is again the relevant operation that defines the word.

The field often adopts the view that words partially contain word-specific heads, like little n (i.e. the nominalizing head). The motivation is almost banal: a morpheme such as a nominalizing affix happens to have a unique role in the structure, so it makes sense to assume a specific functional head at syntax that corresponds to it. In chapter 4 little d was proposed. It is a word-specific head

<sup>192</sup> I am using the word ‘rule’ in a loose sense. I wish to refer to principles such as rules, constraints, parameters, etc., i.e. principles for which it has been argued that they are innate rather than stored.

<sup>193</sup> Words refer here to morphological words, which can contain several phonological words.

within a compound that is reminiscent of a familiar syntactic head (D). Little d has a similar role as D (they both contribute the semantics of referentiality), but they are not identical. Little d does not define an argument in the sentence. It has been observed before that word-internal heads may have specific behavior that deviates slightly from more familiar syntactic heads. For example, it has been noted that the semantics of plurality in the nominal domain or aspect in the verbal domain may show quirky behavior (Acquaviva 2008, Borer 2013a). I think there is still potential in the field to explore such empirical domains and their theoretical consequences (see also De Belder, Faust and Lampitelli 2014): are there principled differences between heads that can occur word-internally and those that cannot?

### 12.3 Morphosyntax and Vocabulary

The present bundle adopts and supports the root hypothesis, according to which the morphosyntactic structure defines the category and the subcategory of the roots (i.e. lexemes) in the structure. In other words, it is denied that lexemes project a categorial superstructure in syntax.

Chapter 2 contributes to the root hypothesis by arguing that bare roots can be observed in the non-head of a specific subtype of primary compounding in Dutch (and other Germanic languages).

Chapter 6 distinguishes between flexibility of the root and grammaticalization. Roots may show malleability in the sense that they can co-occur with functional superstructure of various subcategories. For example, a single root can occur both as an agentive verb and as a stative verb, as argued in the chapter. The occurrence as a functional vocabulary item, however, is not a possibility for a root. As soon as a root occurs as a function word, one has to assume grammaticalization. Malleability is a general possibility in languages and we have seen that the malleability of posture verbs and motion verbs can be observed across the board in continental West-Germanic languages. Grammaticalisation, however, is a language-specific phenomenon and we have seen that there is variation in which verbs had been grammaticalized into a copula.

It became clear throughout the chapters that the ontological nature of the root deserves discussion. The four main proposals in the field are that roots are defined by a root feature (Halle & Marantz 1993, 1994), that roots are defined by means of a phonological index (Borer 2013a), that roots are defined by means of an abstract index (Harley 2014) or that roots are defined by the absence of any features (or indices) on the node (De Belder & Van Craenenbroeck 2015). Harley's proposal seems to be the dominant approach in current work in Distributed Morphology.

The assumption of root indices in the syntactic structure implies that the indices are present in the Numeration. One may then wonder how they were selected into the Numeration. The Lexicon (i.e. the repository of innate features) feeds the Numeration (and thus Syntax). Does the Lexicon contain root indices alongside innate features? Or is the Numeration fed both by the Lexicon and Vocabulary, and if Vocabulary can feed the Numeration, why would it only feed the root's index to the Numeration and not other acquired information? When ones assumes an empty root node or a root feature, one does not have this problem.

The empty node proposal remains the most minimal proposal. As De Belder & Van Craenenbroeck (2015) argue the creation of the root node follows immediately from the operation Merge. The proposal implies that the root cannot project: it is odd to assume that an empty node could define a label or could move. It became clear in chapters 4 and 5 that the proposal might be too restricted. There is an unresolvable contradiction between the assumption that compounding is defined by head movement and the assumption that the head cannot move. Chapter 2 shows a

way out of the puzzle by illustrating how compounding can be derived through fission, i.e. without incorporation, but I admit that I am not sure the proposal can be maintained across the board for compounding.

Given all these considerations, the original idea in the seminal papers of Distributed Morphology that the root node is defined by a root feature has stood the test of time very well, in my opinion. Conceptually, the root feature captures the intuition of the root hypothesis. It expresses that all acquired information is absent from syntax and that the root can simply be defined as a type of node. Given that there is a feature, the node can project and move. The fact that the node is defined by a feature and not by a specific vocabulary item (e.g. the index of a root) is an important advantage of the proposal. It was argued empirically in De Belder & Van Craenenbroeck (2015) that the root node should be defined structurally and not by means of the insertion of a specific vocabulary item. I still stand by that claim.

No-index proposals imply a direct line between PF and the conceptual interface. The conceptual interface depends on input from Vocabulary Insertion to gather information on which roots are inserted in the structure. As such, one's definition of the root node has consequences for one's view on the structure of the Y-model. It is my opinion that such a direct line remains an interesting idea: it connects the acquired *signifié* to the acquired *signifiant*. In such a model, syntax can operate solely on abstract, possibly innate, features and all acquired information is inserted post-syntactically and, indeed, the form-side and the meaning-side of the symbol interface post-syntactically as well.

#### **12.4 Morphosyntax and Semantics**

The work supports the view that functional structure builds meaning step by step in a fully compositional, monotonous, fully specified fashion. Syntactic features have a clear semantic definition and define the category, subcategory and semantics of the underspecified root node. This is exemplified in chapter 6 for verbal functional structure. The semantics of functional features can never be overridden, not even by idiomatic meaning. This holds true for sentences (McGinnis 2002) and words (see for example chapter 4). As such, the present work concurs with most other work in Distributed Morphology.

Chapter 7 describes an alternative view according to which not the root is the most underspecified node, but the transitivising head is (Wood 2015, Wood & Marantz 2016). It is clear that my work does not adopt this approach. It remains a central aspect in my work that the root plays no role of significance at syntax and that syntactic meaning is derived from the functional superstructure.

#### **12.5 Morphosyntax and Phonology**

When it comes to the morphosyntax-phonology interface, I believe that our present understanding of the interfaces fails to capture the essence of the interaction. Both for epenthetic consonants and the relation between syllable structure and categories, we have seen that syntactic structure and phonological constraints can co-determine the licitness and surface form of a syntactic structure. The view that syntax spells out cycles to PF may be too simple. It seems that phonology can also give feedback to syntax, as if they work in loops or as if syntax provides various alternatives to phonology to choose from.

For syntactic alternations such as near-synonym possessive constructions (Rosenbach 2003) or the dative alternation (Bresnan 2007), it has been shown convincingly that semantic factors determine which syntactic structure will be used. It must be that either pre-syntactically semantics

can determine syntax or that post-syntactically, the conceptual interface can crash a less successful syntactic derivation. Similarly, the phonology of a sentence seems to be related to the probability of a syntactic structure as well. It has, for example, been shown that the rhythm of the sentence correlates with the chosen word-order in the Dutch verbal group (De Sutter, Speelman & Geeraerts 2008). Chater 11 in this bundle shows that syllable structure correlates with categories. There must be a way to capture such observations. How exactly they may happen (pre-syntactically/post-syntactically, at the end of the complete derivation/ cyclically, etc.) is at this point an open question. Clearly, the model and the technicalities need to be developed. But note that interactions between morphosyntactic structure and phonological structure (Wolf 2008) are then expected: they result from an architectural, deterministic relation between syntactic and phonological structure (pace Wolf 2016). It is a problem for the current model of Distributed Morphology that such probability-based observations cannot be captured.

## 12.6 Notes on compounding

The present proposal identified functional heads within the Dutch primary compound. It was argued that the non-head can be a bare root, that it can contain nominal class features or a little d head. I have also described West-Flemish compounds in which the non-head clearly surfaces with a gender head. The generalization is rather clear: the non-head in the Dutch compound seems either a bare root or it contains nominal projections. This empirical domain is underexplored and barely understood. The study as to which heads can be contained in the primary compound's non-head in West-Germanic languages deserves a principled approach, parallel to the exploration of functional heads in the DP we have seen thirty years ago. I hope to explore this field in more detail at a later point in my career.

## 12.7 Notes on nominal structure

There is something about nouns. They are different than other open-class lexical items. They seem to be 'privileged'. Firstly, nominal structure seems to contain structure that is neither fully functional or lexical, such as noun classes, gender heads and number heads (see also Aquaviva 2008 and see Borer 2013a on nominal plurality as not fully functional). These heads seem to be selected by the specific vocabulary item in the root and they seem to form stored combinations of meaning with the root. It is, for example, well-known, that in languages with syntactic gender, the gender of a noun 'has to be learned'. Secondly, it becomes clear in this dissertation that the non-head of compounds are roots, nouns or phrases. In other words, the only non-heads for which one finds evidence that a root has merged with functional structure in the non-head without being part of a larger phrases are nominal non-heads. Thirdly, it also becomes clear that nouns can have more syllables and more complex syllables than any other category in Dutch. These observations are not unique to Dutch: class, gender and number projections, nominal non-heads in compounds and prosodic nominal privilege are all known cross-linguistically. I have to admit I do not have an answer as to why nouns should be special. Perhaps they are not and the list of observations do not hang together in a principled fashion, but perhaps nominal structure does differ from other open-class structures.

# Zusammenfassung

## 1. Forschungsthemen

Die vorliegende Arbeit diskutiert die niederländische Morphologie und ihre Schnittstellen. Morphologie hat verschiedene Schnittstellenbeziehungen. Erstens hat die Morphologie natürlich eine Schnittstellenbeziehung zum Wortschatz der Sprache. Morphologie baut Wörter und die Morpheme des Wortes werden durch Wortschatzelemente (d. h. Wurzeln und Affixe) realisiert. Zweitens muss angesichts der offensichtlichen Tatsache, dass Sätze aus Wörtern bestehen, eine enge Schnittstellenbeziehung zwischen dem Modul, das Wörter bildet, und dem Modul, das Sätze bildet, bestehen. Drittens ist die abstrakte Ausgabe der Morphologie die Eingabe der Wortphonologie. Viertens bedeuten Wörter etwas. Sie können lexikalische Semantik und funktionale Semantik zum Satz beitragen. Tatsächlich können sie sogar mit Pragmatik verbunden sein, wenn sie beispielsweise ehrenhafte Affixe enthalten.

Der Linguist oder die Linguistin, der Morphologie und ihre Schnittstellen studiert, steht vor zwei wichtigen Herausforderungen. Erstens werden Daten nicht beschriftet, daher besteht das erste Problem oft darin, zu bestimmen, für welches Modul oder welche Schnittstelle zwischen Modulen das empirische Problem relevant sein könnte. Ist zum Beispiel die Allomorphie-Auswahl ein striktes Problem des Wortschatzes (d.h. der Wortschatz hat mehrere Variationen eines Exponenten und muss sie organisieren), ist sie ein Effekt der Art und Weise, wie die morphosyntaktische Struktur mit dem Wortschatz verbunden ist (d.h. einer der Allomorphe kann markiert sein, um eine spezifische morphosyntaktische Struktur zu realisieren) oder ist es ein phonologisches Problem (z.B. kann ein Allomorph im Kontext einfach phonologisch optimaler sein). Die erste Forschungsfrage des vorliegenden Bündels ist daher die Frage, welches Modul bzw. welche Schnittstelle der adäquate Ort für bestimmte scheinbar hybride Daten ist. Die Analyse solcher Daten wird seit Jahrzehnten oft diskutiert. Die vorliegende Dissertation stellt beispielsweise Verbindungen vor, die eindeutig eine syntaktische Phrase oder ein Fragment einer syntaktischen funktionalen Struktur enthalten. Man könnte sich natürlich fragen, was uns solche Verbindungen über die Verbindungsbildung lehren. Allgemeiner könnte man jedoch fragen, was wir aus solchen Daten über die Beziehung zwischen Morphologie und Syntax lernen. Ein weiteres Beispiel ist die syntaktische Polysemie. Wenn ein Verb sowohl als lexikalisches Verb als auch als Kopula verwendet werden kann, ist dies dann eine Komplikation in der Syntax, im Wortschatz oder vielleicht beides? Was sagen uns solche Daten über diese Module und ihre Schnittstellen?

Die zweite Herausforderung besteht darin, dass die Daten normalerweise keine sauber geordnete Interaktion zwischen zwei Modulen zeigen. Stattdessen scheinen alle Module gleichzeitig miteinander zu interagieren. Betrachten Sie das folgende Beispiel. Es wird deutlich, dass das Einfügen eines bestimmten epenthetischen Konsonanten im Niederländischen durch das gleichzeitige Auftreten eines bestimmten morphosyntaktischen, lexikalischen und phonologischen Kontexts bestimmt und durch einen ehrenhaften (d. h. einen bestimmten pragmatischen) Kontext blockiert wird. Theoretisch ist ein solcher Zustand nicht zu erwarten. Man erwartet eher ordentliche Output-Input-Schnittstelleninteraktionen zu finden. Ich habe jedoch festgestellt, dass dies empirisch selten der Fall ist. Es besteht somit eine Spannung zwischen der Komplexität, die die Daten suggerieren, und der Tatsache, dass wir vernünftigerweise davon ausgehen können, dass das Modell ein streng organisierter Prozess sein muss. Das zweite Thema der Dissertation ist daher die

Frage, wie das Y-Modell und insbesondere die Schnittstellen im Y-Modell zur Erfassung der Daten zu organisieren sind. Dabei berücksichtige ich theoretische Überlegungen, die der Frage nachgehen, was wir aus prinzipiellen und technischen Erwägungen unter einer wünschenswerten Organisation erwarten, und garantiere, dass wir am Ende des Tages über die Werkzeuge verfügen, um die Oberflächendarstellung abzuleiten.

Ich behandle damit beide Herausforderungen. Ich präsentiere empirische Domänen, die für unser Verständnis relevant sind, welches Modul und welche Schnittstellen für welches Phänomen relevant sind und diskutiere theoretische Überlegungen und technische Lösungen, um ein adäquates Schnittstellenmodell zu organisieren. Fast alle Beiträge diskutieren die niederländische Morphosyntax. Zwei Ausnahmen sind der kurze Beitrag in Kapitel 9 zum Bretonischen und der allgemeinere theoretische Beitrag in Kapitel 7, der relevante Beispiele aus verschiedenen Sprachen diskutiert.

Die vorliegende Arbeit ist eingebettet in den Rahmen der Distribuierte Morphologie. Die Struktur ihres Y-Modells hat unmittelbare Konsequenzen für mehrere Schnittstellen. Erstens geht das Framework nur von einem einzigen strukturerzeugenden Modul namens Syntax aus. Die Erzeugung von Wörtern und Sätzen findet nämlich im selben Modul statt. Mit anderen Worten, Morphologie ist Syntax und ich werde daher alle Strukturen, Wörter oder Sätze in dieser Arbeit morphosyntaktische Struktur nennen. Es ist daher technisch ungenau, in diesem Rahmen über die Morphologie-Syntax-Schnittstelle zu sprechen: Ein Modul verbindet sich nicht mit sich selbst. Dies macht die Beziehung zwischen syntaktischer Struktur und morphologischer Struktur jedoch nicht weniger interessant, wie in einigen Kapiteln deutlich wird. Jeder Morphologe und jede Morphologin muss die Beziehung und die Kluft zwischen Satzstruktur und Struktur, die die Worthaftigkeit definiert, verstehen, unabhängig davon, wie viele Module an der Darstellung beteiligt sind.

Zweitens ist die Beziehung zwischen Syntax und Vokabular in der Distribuierten Morphologie stark durch die wurzelbasierte Natur des Frameworks gekennzeichnet. Das Framework nimmt keine kategorialen Merkmale für lexikalische Elemente der offenen Klasse im Lexikon an. Diese Hypothese hat weitreichende Konsequenzen für die Vokabular-Morphosyntax-Schnittstelle (und die Semantik-Morphosyntax-Schnittstelle). Die vorliegende Arbeit trägt zum Verständnis dieser Konsequenzen bei.

Drittens gibt es in der Distribuierten Morphologie keinen Konsens darüber, wie Morphologie und Phonologie im Modul zusammenwirken. Da das Modul die allgemeine Y-Struktur übernimmt, wird die grundsätzliche Reihenfolge der Schnittstellen vereinbart. Die spezifischen technischen Details der Interaktion zwischen Morphologie und Phonologie werden jedoch noch diskutiert. Die vorliegende Arbeit ergänzt diese Debatte durch die Diskussion von Fallstudien, die zu unserem Verständnis der Komplexität der Daten beitragen, und bietet theoretische Vorschläge und technische Lösungen.

Viertens ist die Schnittstelle zwischen Morphologie und Semantik selten das Hauptthema der Arbeit in der Distribuierten Morphologie, aber angesichts der spezifischen Sichtweise des Fachs auf das Vokabular interessant. Wenn für Kategorien oder andere Merkmale keine Stammvokabularelemente angegeben sind, woher kommt dann die kategoriale Bedeutung? Verschiedene Vorschläge haben zu sehr unterschiedlichen Annahmen über die morphosyntaktische Struktur geführt. Die vorliegende Arbeit enthält einen kritischen Überblick über den Stand des Frameworks in dieser Hinsicht und bietet auch Vorschläge.

Kurz gesagt, die vorliegende Dissertation ist eingebettet in die Distribuierten Morphology. Die Konzeption des Modells in diesem Rahmen definiert unmittelbar, wie bestimmte Schnittstellen zu

verstehen sind, bestimmt aber auch die offenen Fragen. Die vorliegenden Vorschläge tragen zu diesem Rahmen bei.

## 2. Kapitel

### 2.1 Komposita (Kapitel 2 bis 5)

Die Arbeit enthält drei Artikel zu primären Komposita im Niederländischen und einen Artikel zu Phrasal Compounding. Die primäre Komposita sind die „grundlegende“ Art der Komposita Compoundierung: Der Nicht-Kopf enthält ist keine Phrase und der rechte Teil ist der Kopf. Ich argumentiere in diesen Artikeln, dass im Niederländischen (mindestens) drei Subtypen der primären Komposita unterschieden werden können. Sie sind in (22) dargestellt.

(1) a. kreeft-Woord	b. kreeft-en-soep	c. Kreeft-s-keerkring
Hummer-Wort	Hummer-DE-Suppe	Hummer-S-tropisch
‘Palindrom’	‘Hummersuppe’	‘Wendekreis des Krebses’

Ich argumentiere, dass der in (1)a dargestellte Subtyp eine bloße Wurzel als Nicht-Kopf enthält. Empirisch führt die bloße Wurzel dazu, dass es kein sogenanntes „verknüpfendes Element“ gibt. Immerhin ist die Wurzel kahl und mit nichts verschmolzen, was nachträglich durch ein Affix offen realisiert werden könnte. Darüber hinaus wird der Nicht-Kopf nicht unbedingt als nominal interpretiert (auch wenn dies in diesem Beispiel der Fall wäre). Für den Subtyp in (1)b argumentiere ich, dass der Nicht-Kopf einen nominellen funktionellen Kopf enthält, genauer gesagt eine nominelle Klassenprojektion. Die nominelle funktionale Projektion wird entweder als *-s-* oder *-en-* realisiert, abhängig von der spezifischen Wurzel im Nicht-Kopf (für *kreeft* "Hummer" wird *-en-* gewählt, wie im Beispiel zu sehen ist). Selbstverständlich definiert der funktionale Kopf den Nicht-Kopf als nominell und wird als solcher interpretiert. Für den Subtyp in (1)c argumentiere ich, dass der Nicht-Kopf einen Determinator enthält, der als *-s-* auftaucht. Dies führt dazu, dass der Nicht-Kopf referentielle Eigenschaften hat. Auch dieser Nicht-Kopf wird aufgrund der nominellen Funktionsprojektion als inhärent nominell interpretiert. Der Vorschlag, dass der Nicht-Kopf referentiell sein kann, ist eine ziemlich neue Behauptung: Es wurde angenommen, dass referentielle Nicht-Köpfe grundsätzlich aus Verbindungen in jeder Sprache ausgeschlossen sind. Allgemeiner ausgedrückt sind die Identifizierung dieser drei Subtypen für die niederländische primäre Komposita und die von mir bereitgestellten Analysen neue Ansätze für die niederländische Komposita, auch wenn ich zugegebenermaßen gar nicht der Erste bin, der vermutet, dass (1)a und (1)b abgesondert werden (siehe Bilderdijk 1826:207, der bereits eine kurze Bemerkung formuliert hat).

Kapitel 2 über die Wurzelprimärverbindung umfasst die Morphosyntax-Lexikon-Schnittstelle mit Konsequenzen für die Semantik. Der Nicht-Kopf dieser Art der Komposita ist völlig akategorial. Die enthalten keine kategorialen Köpfe (d. h. kleine Köpfe) oder funktionale Projektionen. Interpretationstechnisch bleibt die Kategorie des Nicht-Kopfes syntaktisch unaufgelöst (und hängt vom zu interpretierenden Weltwissen ab). In einigen Fällen kann dies tatsächlich zu Mehrdeutigkeiten führen:

(2) Antwoord-centrum

Antwort-Zentrum

'Zentrum, wo Antworten gespeichert werden / Zentrum, wo Antworten bereitgestellt werden/...' (Nennwerte)

'Zentrum, wo man ans Telefon geht/...' (verbales Vorlesen)

Die Diskussion unterstützt die Hypothese, dass die minimalste lexikalische Projektion eine Wurzel ist (Halle und Marantz 1993, Borer 2005a) und dass sie einen syntaktischen Slot in der Struktur einnimmt, der nicht von der Überstruktur lizenziert werden muss, um semantisch interpretierbar oder realisierbar zu sein. Dies ist eine wichtige Erkenntnis, da davon ausgegangen wurde, dass nackte Wurzeln grundsätzlich ausgeschlossen werden, da sie nicht interpretierbar wären (z. B. Harley 2012, Starke 2009 u.a.). Diese Verbindungen widerlegen diese Annahme und zeigen, dass die kategoriale Zuordnung keine zwingende Voraussetzung für die Interpretation ist. Zugegeben, das Endergebnis kann mehrdeutig sein, aber in seiner Mehrdeutigkeit ist es perfekt interpretierbar: Die Struktur hat am Ende einfach mehrere Interpretationen.

Der Artikel enthält auch Überlegungen, wie man morphologische Komplexität überhaupt erkennen kann. Dieses Problem wird meiner Meinung nach oft übersehen. Um ein korrektes Verständnis der Vokabular-Morphosyntax-Schnittstelle zu erreichen, ist es wichtig, die relevanten Daten zu erkennen. Es ist unnötig zu erwähnen, dass morphologisch simplexe Wörter für eine solche Studie nicht relevant sind. Der Artikel argumentiert auch für eine direkte Beziehung zwischen PF und LF. Es wird deutlich, dass bei der morphosyntaktischen Ableitung kein Wurzelindex angenommen wird, sondern eine underspezifizierte Wurzel, und dass Wurzeln nur bei der Vokabulareinfügung eingefügt werden. Eine solche Analyse erfordert eine direkte Verbindung zwischen PF und der konzeptionellen Schnittstelle, sonst hätte die konzeptionelle Schnittstelle keinen Zugriff auf die Inhalte der Wurzeln.

**Kapitel 3** behandelt sogenannte Linking-Elemente. Sprachübergreifend können sie tatsächlich zwei sehr unterschiedliche Rollen erfüllen: Sie können phonotaktische Elemente sein oder sie können einen morphosyntaktischen Kopf realisieren. Als solche handelt es sich um Daten, bei denen *a priori* nicht klar ist, für welches Modul oder welche Schnittstelle sie überhaupt relevant sind. Ich präsentiere eine Fallstudie über verbindende Elemente im Niederländischen, die eine morphosyntaktische Rolle erfüllen. Ich stelle zunächst fest, dass sich der Linguist oder die Linguistin der Tatsache bewusst sein sollte, dass es sich bei diesen „Verknüpfungselementen“ um Nullmorpheme handeln kann. Ich zeige, dass das Niederländische tatsächlich ein solches Null-Affix in seinem Inventar haben kann. Ich argumentiere dann, dass die niederländischen Verknüpfungselemente in Wirklichkeit Nomenklassenmarkierungen sind, und leite daher die Tatsache ab, dass zwei von ihnen durch kumulative Exponenten mit der Pluralmarkierung identisch sind. Ich weise darauf hin, dass diese Substantivklassen mit der Geschlechtskennzeichnung koexistieren und sogenannte ‘fehlerhafte’ Verweise verursachen können.

Die Analyse zeigt, dass uns eine sorgfältige Untersuchung solcher Elemente gleichzeitig neue Einsichten in zwei Domänen liefert. Zuerst erfahren wir mehr über die Morphosyntax des linken Teils der Verbindung, die Struktur der Verbindung selbst und das Inventar der Verbindungstypen in der Sprache. Die Bedeutung des linken Teils ist meiner Meinung nach nicht zu unterschätzen. Es ist eine einzigartige Domäne: Sie zeigt, wie sich eine Kategorie funktional verhält, wenn sie nicht als Argument fungiert. Als solche erfahren wir zweitens etwas über die

Morphosyntax einer Kategorie. In diesem Beitrag scheint es, dass die niederländische Nominaldomäne Nominalklassen hat, die in der Nominalgruppe weniger sichtbar sind, da sie mit der Pluralaffixen synkretistisch sind. Nur in Verbindungen ohne Pluralaffixen werden sie sichtbar. Da die Klassenmarkierung von Substantiven eine syntaktische Projektion ist, wird argumentiert, dass eine syntaktische Projektion in der Verbindung enthalten ist. Unter der Annahme, dass Morphologie Syntax ist, wird ein solcher Sachverhalt erwartet. Es wird deutlich, dass es eine Selektionsbeziehung zwischen einer Wurzel und einer Substantivklasse geben scheint, genauso wie es eine Selektionsbeziehung zwischen einer Wurzel und einer Geschlechtsklasse geben kann. Der Vorschlag ist, dass die Verbindung bei *Encyclopedia* (eine Liste für lexicalische Semantik im Modell der Distriuierten Morphologie) gespeichert wird. Syntax kann alles generieren, aber nicht alles ist am Ende des Tages interpretierbar. Nur syntaktische Strukturen, die mit gespeicherten Interpretationen übereinstimmen (z. B. aufgelistete Kombinationen einer Wurzel und ihres Geschlechts oder ihrer Klasse) sind interpretierbar.

**Kapitel 4** enthält eine weitere Studie über nominale Projektionen in Komposita. Es wird allgemein angenommen, dass Nicht-Kopfzeilen von Komposita nicht referentiell sein können: Es wird angenommen, dass ihnen eine D-Schicht fehlt. In diesem Artikel argumentiere ich, dass diese Annahme falsch ist. Das Niederländisch hat Komposita mit einem referentiellen Nicht-Kopf. Diese Nicht-Köpfe gehören zu den Arten von Lexemen, die Longobardi (1994) als typischerweise einer N-nach-D-Erhebung unterworfen identifizierte: Eigennamen, Verwandtschaftsnamen, Namen von Wochentagen und Feiertagen, einzigartige Entitäten und kontextuell einzigartige Entitäten. Ich argumentiere, dass die Komposita einen Nominalisierungskopf enthalten, der für Einzigartigkeit und Referenzialität spezifiziert ist, was ich kleines d nenne. Der Nicht-Kopf der Verbindung gliedert sich in kleine d ein. Ich vergleiche die Morphosyntax dieser Nicht-Köpfe mit der Morphosyntax von modifizierenden Verbindungen, von Eigennamen in der Argumentposition, der niederländischen Possessiv- und Titelausdrücke, die meiner Meinung nach niederländische referentielle Konstruktzustände sind.

Zusammen mit Kapitel 2 und Kapitel 3 untersucht die Arbeit die Art von Projektionen, die auf dem Nicht-Kopf der Verbindung erlaubt sind. Diese Domäne ist noch ziemlich unerforscht und ich möchte darauf hinweisen, dass in dieser Domäne mehrere nominelle Köpfe enthalten sein können und deren Vorhandensein oder Fehlen mit relevanten semantischen Unterscheidungen einhergeht. Die Semantik von Komposita wird daher nicht einfach aufgelistet: Sie zeigt vorhersagbare kompositorische semantische Muster. Dies erinnert an McGinnis' (2002) Arbeit über Idiome: Semantische Bedeutung kann gespeichert werden, aber strukturelle Bedeutung kann nicht außer Kraft gesetzt werden. Eine nackte Wurzel ist kategorial mehrdeutig, eine nominale Klassenprojektion führt zu einer modifizierenden Verbindung und eine referentielle Schicht führt zu einer referenziellen Verbindung.

Kapitel 5 ist das letzte Kapitel über Komposita. Den drei Artikeln zum primären Compounding folgt ein Beitrag zum ANN-Compounding im Niederländischen. In diesem Beitrag liefern Marjo Van Koppen und ich empirische Beweise gegen die Behauptung, dass die Morphologie im Gegensatz zur Syntax im Umgang mit Elementen steht, die im Lexikon aufgeführt sind. (Jackendoff 1975, Aronoff 1976, Jensen und Stong-Jensen 1984). Genauer gesagt unterscheiden wir im Niederländischen zwischen drei Arten von ANN-Verbindungen. Wir zeigen, dass die strukturellen Eigenschaften dieser Typen keine Eins-zu-Eins-Abbildung mit lexikalischen Eigenschaften aufweisen, wie beispielsweise eine aufgelistete oder sogar idiomatische Bedeutung (siehe DiSciullo & Williams 1987). Auf dieser Grundlage argumentieren wir, dass

Schlussfolgerungen über die Struktur bestimmter morphologisch komplexer Wortformen auf strukturellen Eigenschaften und nicht auf lexikalischen Eigenschaften wie Idiomatik oder Lexikalisierung basiert sein sollten. Wir schlagen eine syntaktische Ableitung für alle Arten von ANN-Verbindungen im Niederländischen vor (pace Ackema und Neeleman 2004). Strukturelle Unterschiede ergeben sich aus der Ebene der Verschmelzung: Was wir traditionell Morphologie nennen, ist die Syntax unterhalb des funktionalen Bereichs.

## 2.2 Haltungsverben, Bewegungsverben und Verknüpfungsverben (Kapitel 6)

In der typologischen Literatur wurde beobachtet, dass Haltungsverben oft zu Verknüpfungsverben grammatisieren (Newman 2002). Artikel 6 stellt dieses Phänomen dar, indem er das Phänomen bei niederländischen und anderen kontinentalen westgermanischen Varietäten diskutiert, d.h. Hochdeutsch, Plattdeutsch und Friesisch. Genauer gesagt kontrastiert er Formbarkeit mit Grammatikalisierung im empirischen Bereich der Haltungsverben und Bewegungsverben. Diese Verben können alle als lexikalische Verben verwendet werden. Sie zeigen aber auch gebleichte Verwendungen, die wie Verknüpfungsverben aussehen. Bei einigen Verben ergibt sich diese Ähnlichkeit mit Verknüpfungsverben lediglich aus der Tatsache, dass sie in nicht-agentiven, statischen Kontexten vorkommen können, was im Gegensatz zu der handelnden Aktivität des Lesens in ihrer lexikalischen Verwendung steht. Ich schließe daraus, dass diese Aspektmerkmale daher nicht von Natur aus zum Verb gehören, sondern Formbarkeit zeigen. Andere Verben wurden jedoch tatsächlich in Verknüpfungsverben grammatisiert. Der Artikel untersucht daher die Schnittstelle zwischen Syntax und Wortschatz: Er unterscheidet zwischen Polysemie durch Grammatikalisierung und Speicherung eines funktionalen Wortschatzelements im Wortschatz und der bloßen Formbarkeit von Wortschatzelementen. Er trägt auch zu der Ansicht bei, nach der Semantik kompositorisch durch Merkmale in den morphosyntaktischen Strukturen aufgebaut wird (siehe auch Kapitel 7).

## 2.3 Die Syntax-Semantik-Schnittstelle in der Distribuierten Morphologie (Kapitel 7)

Kapitel 7 gibt einen Überblick darüber, wie die Syntax-Semantik-Schnittstelle in der Distribuierten Morphologie verstanden wird. Literatur der Distribuierten Morphologie neigt dazu, dem Prinzip der vollständigen Kompositionalität auf der LF-Seite zu folgen: Nur wenige Autoren haben postsyntaktische strukturverändernde Operationen auf der semantischen Seite der Berechnung vorgeschlagen. Das Feld geht jedoch aufgrund der Unterspezifikation der syntaktischen Struktur im Allgemeinen nicht von einer einfachen Übereinstimmung zwischen syntaktischer und semantischer Struktur aus. Die Idee, dass Syntax semantisch unterspezifiziert werden kann, wurde auf verschiedene Weise untersucht. Die Idee geht auf den Wurzelknoten zurück, der seit den bahnbrechenden Arbeiten als unterspezifiziert gilt. Diese Idee verbreitete sich allmählich auf kleine v und allgemeiner auf transitive Köpfe. Daher haben sich einige Autoren von der ursprünglichen Idee entfernt, dass es eine enge, präzise und deterministische Übereinstimmung zwischen den morphosyntaktischen Merkmalen der Struktur und ihrer Bedeutung gibt.

## 2.4 Allomorphie (Kapitel 8 und 9)

Artikel 8 argumentiert, dass die Auswahl von Allomorphen auf zwei Module verteilt ist, nämlich Wortschatzeinfügung und Phonologie. Dies geschieht auf der Grundlage einer Fallstudie zu alternierenden Allomorphen der Vokallänge im Niederländischen. Die Daten zeigen ein

gespaltenes Muster: Einige empirische Bereiche können vollständig durch phonologische Prinzipien erfasst werden. Für diese Fälle wird das phonologisch optimale Allomorph ausgewählt. In anderen empirischen Bereichen machen phonologische Prinzipien noch immer viele der beglaubigten Daten aus. Allerdings attestiert man auch lexikalisierte Ausnahmen, die eindeutig phonologisch nicht optimal sind. Die Daten spiegeln widersprüchliche Ansichten in der Literatur wider: Einige Vorschläge versuchen, die Allomorphselektion auf die Phonologie zu reduzieren, andere konzentrieren sich auf die Tatsache, dass viele Beispiele einfach nicht phonologisch optimal sind und legen nahe, dass die Allomorphselektion nicht durch die Phonologie erfolgen sollte. Ich argumentiere, dass die gegensätzliche Natur dieser beiden Datentypen tatsächlich ein Hinweis darauf ist, wie die Auswahl von Allomorphen organisiert ist. Genauer gesagt können sowohl die Vokabeleinfügung als auch die Phonologie die Auswahl von Allomorphen bestimmen. Die Vokabeleinfügung ist für gespeicherte Informationen verantwortlich, die Phonologie für die phonologische Optimierung von Mustern.

**Artikel 9** zeigt, dass die bretonische Pluralität weitere empirische Unterstützung für die Hypothese liefert, dass die geeignete Lokalitätsbedingung für die Allomorphie die Kopfinkorporation ist. Angesichts der Tatsache, dass die Kopfinkorporation die Wort-Haube definiert, besteht eine enge Verbindung zwischen der Wort-Haube und dem geeigneten Bereich für Allomorphie. Es wurde allgemein angenommen, dass der Auslöser der Allomorphie ausreichend lokal für das Morphem sein sollte, das der Allomorphie unterliegt (Embick 2010). Bobaljik (2012:68) nimmt als Arbeitshypothese an, dass diese Lokalität als Kopfinkorporation definiert werden sollte: Ein Kopf, der die Wurzelallomorphie bedingt, muss sich im selben komplexen Kopf (d. h. demselben morphologischen Wort) wie die Wurzel befinden. Als solches erfasst er die Tatsache, dass die Komparative Supplementierung in peripherastischen Komparativen blockiert ist. Diese Annahme wird in Kapitel 9 empirisch gestützt.

## 2.5 Epenthetische Konsonanten und Silben und Wortarten (Kapitel 10 und 11)

Das Bündel schließt mit zwei Artikeln, die zeigen, dass syntaktische Merkmale phonologische Beschränkungen bestimmen können: Die syntaktische Struktur kann die Einfügung eines epenthetischen Konsonanten oder die Akzeptanz einer Silbenstruktur bestimmen.

**Kapitel 10** untersucht die Einfügung eines epenthetischen Konsonanten, der durch das Affix *-ke* bestimmt wird. Das niederländische Diminutivsuffix *-ke* kommt in Wörtern aus vier verschiedenen Kategorien vor: Substantive, prädikativ verwendete Adjektive, Adverbien und Interjektionen. Bei Substantiven, Adjektiven und Adverbien kann das Suffix einen epenthetischen Konsonanten auslösen, der für die Interjektionen nicht zur Verfügung steht. Damit scheint eine syntaktische Unterscheidung einherzugehen: Bei den Interjektionen ist das Suffix nicht als Diminutiv, sondern als Ehrenzeichen zu analysieren. Ich argumentiere, dass das Suffix in allen Kontexten skalare Merkmale realisiert, jedoch unterschiedlicher Natur. Der epenthetische Konsonant hängt mit dem skalaren Merkmal von der Lokalität ab.

In **Kapitel 11** von mir und Esther Ruigendijk wird die Frage gestellt, ob Trommelen's Beobachtung einer Beziehung zwischen niederländischer Silbenstruktur und syntaktischen Kategorien mit zeitgenössischen experimentellen Werkzeugen bestätigt werden kann. Trommelen (1989) beobachtete, dass niederländische Substantive und Verben sich in der Silbenstruktur zu unterscheiden scheinen, aber diese Hypothese wurde nie konkretisiert. Wenn Trommelen Recht hätte, wäre Niederländisch eine der Sprachen, die suprasegmentale Substantiv-Verb-Unterschiede aufweisen (Smith 2011). Der Artikel liefert tatsächlich Beweise für phonologische Substantiv-Verb-

Unterschiede im Niederländischen (siehe auch Trommelen 1989; Don und Erkelens 2006). Durch korpusbasierte Computerforschung konnten wir bestätigen, dass Verbalstämme weniger Silben und weniger komplexe Silben enthalten als Nominalstämme. In einem groß angelegten Online-Experiment haben wir zudem festgestellt, dass Muttersprachler aufgrund ihrer Silbenstruktur eine Intuition für die Wortklasse eines Pseudowortes haben. Auch hier werden die Intuitionen durch die Komplexität der Silben und die Anzahl der Silben gesteuert. Wichtig ist, dass die experimentellen Ergebnisse etwas andere Muster aufweisen als die Korpusforschung. Dies deutet darauf hin, dass etwas die Intuition in der sprachlichen Kompetenz steuert, das über das bloße Bewusstsein statistischer Muster im Lexikon hinausgeht. Diese Ergebnisse werden im Lichte des Optimal Paradigms-Ansatzes diskutiert, der eine Erklärung für ähnliche Tatsachen in anderen Sprachen vorschlägt, indem er auf Flexionsunterschiede zwischen Verben und Substantiven hinweist (McCarthy 2005; Cable 2004). Basierend auf unseren Ergebnissen können Flexionsunterschiede die niederländischen Fakten nicht erklären und wir haben die Hypothese falsifiziert, dass das Paradigma als Ganzes in einem Experiment Druck auf die Form des Stammes ausüben kann. Als solche unterstützt die Studie Bobaljiks (2006) Skepsis bezüglich der Bezugnahme auf Paradigmen bei der Berücksichtigung phonologischer Nomen-Verb-Unterschiede. Wir haben auch deutsche Konversionsdaten, die zeigen, dass die relevante Schnittstellenbeziehung zum Verständnis dieser Domäne nicht zwischen Vokabular und Syntax zu verorten ist, sondern zwischen Syntax und Phonologie.

## References

- Abney, Steven Paul (1987) *The English noun phrase in its sentential aspect*. Doctoral dissertation, MIT.
- Ackema Peter & Ad Neeleman (2004) *Beyond Morphology*. Oxford: Oxford University Press.
- Acquaviva, Paolo (2008) *Lexical Plurals: A Morphosemantic Approach*. Oxford: Oxford University Press.
- Aikhenvald, Alexandra (2000) *Classifiers: a typology of noun categorization devices*. Oxford: Oxford University Press.
- Alexiadou, Artemis & Elena Anagnostopoulou (2008) Structuring participles. In: Charles B. Chang and Hannah J. Haynie (eds.), *Proceedings of the 26th West Coast Conference on Formal Linguistics*. Somerville, MA: CascadillaPress.
- Alexiadou, Artemis, Elena Anagnostopoulou, and Florian Schäfer (2015) *External Arguments in Transitivity Alternations: A Layering Approach*. Oxford Studies in Theoretical Linguistics 55. Oxford: Oxford University Press.
- Alexiadou, Artemis, Hagit Borer, and Florian Schäfer (eds.) (2014) *The syntax of roots and the roots of syntax*. Oxford: Oxford University Press.
- Alexiadou, Artemis, Liliane Haegeman and Melita Stavrou (2007) Noun phrase in the generative perspective. *Studies in Generative Grammar Volume 71*. Berlin & New York: De Gruyter Mouton.
- Anderson, Stephen R. (1986) Disjunctive ordering in inflectional morphology. *NLLT* 4. 1-31.
- Anderson, Stephen R. (2008) Phonologically conditioned allomorphy in Surmiran (Rumantsch). *Word Structure* 1. 109–134. DOI: 10.1093/acprof:oso/9780199589982.003.0002
- Anttila, Arto (2002) Morphologically conditioned phonological alternations. *NLLT* 20. 1-42.
- Arad, Maya (2005) *Roots and Patterns. Hebrew Morpho-syntax. Studies in NLLT volume 63*. Dordrecht: Springer.
- Audring, Jenny (2009) *Reinventing Pronoun Gender*. Utrecht: LOT Dissertation Series.
- Bailyn, John F. and Andrew Nevins (2008) Russian genitive plurals are impostors. In: Bachrach, Asraf and Andrew Nevins (eds.) *Inflectional Identity*, 237-270. Oxford: Oxford University Press.
- Baker, M. (1985) The mirror principle and morphosyntactic explanation. *Linguistic Inquiry* 16(3). 373-415.
- Bat-El, Outi (1994) Stem modification and cluster transfer in Modern Hebrew. *NLLT* 20. 571-596.
- Bauer, Laurie (1978) *The grammar of nominal compounding*. Odense: Odense University Press.
- Beard, Robert (2001) Derivation. In: Spencer, Andrew and Arnold M. Zwicky (eds.) *The handbook of morphology*, 44-65. Hoboken: John Wiley and Sons.
- Bilderdijk, Willem (1826) *Nederlandsche Spraakleer*. Den Haag: J. Immerzeel Junior.
- Bobaljik, Jonathan D. (1994) What does adjacency do? In: Harley, H. & C. Phillips (eds.) *MITWPL 22: The Morphology-Syntax Connection*, 1-32. Cambridge, MA: MIT Working Papers in Linguistics.
- Bobaljik, Jonathan D. (1998) Mostly predictable: cyclicity and the distribution of schwa in Itelmen. Samiian Vida (ed.) *Proceedings of WECOL 1996*, 14-28. Fresno: U.C. Santa Cruz.
- Bobaljik, Jonathan D. (2002) Syncretism without paradigms: remarks on Williams 1981, 1994. In: Booij, Geert and Jaap van Marle (eds.) *Yearbook of Morphology 2001*, 53-85. Dordrecht: Foris.
- Bobaljik, Jonathan D. (2006) Paradigms (optimal and otherwise). A case for scepticism. In: Bachrach, A. & A. Nevins (eds.) *Inflectional Identity*, 29-54. Oxford: OUP.

- Bobaljik, Jonathan D. (2012) *Universals in Comparative Morphology. Suppletion, Superlatives, and the Structure of Words*. Cambridge, MA & London: The MIT Press.
- Bobaljik, Jonathan D. & Höskuldur Thráinsson (1998) Two heads aren't always better than one. *Syntax* 1. 37-71
- Bonet, Eulàlia (1991) *Morphology after Syntax: Pronominal Clitics in Romance*. Doctoral dissertation, MIT.
- Bonet, Eulàlia & Daniel Harbour (2012) Contextual allomorphy. In: Jochen Trommer (ed.) *The morphology and phonology of exponence*, 195-235. Oxford: Oxford University Press.
- Booij, Geert (1995) *The phonology of Dutch*. Oxford: Oxford University Press.
- Booij, Geert (1998) Prosodic output constraint in morphology. In: Kehrein, Wolfgang and Richard Wiese (eds.) *Phonology and morphology of the Germanic languages*, 143-163. Tübingen: Niemeyer.
- Booij, Geert (1999) *The phonology of Dutch*. Oxford: Oxford University Press.
- Booij, Geert (2002) *The morphology of Dutch*. Oxford: Oxford University Press.
- Booij, Geert (2010) *Construction Morphology*. Oxford: Oxford University Press.
- Borer, Hagit (1984) *Parametric Syntax. Studies in Generative Grammar* 13. Dordrecht: Foris Publications.
- Borer, Hagit (2001) Morphology and syntax. In: Spencer, Andrew and Arnold M. Zwicky (eds.) *The handbook of morphology*, 44-65. Hoboken: John Wiley and Sons.
- Borer, Hagit (2005a) *In name only. Structuring Sense Volume I*. Oxford: Oxford University Press.
- Borer, Hagit (2005b) *The Normal Course of Events. Structuring Sense Volume II*. Oxford: Oxford University Press.
- Borer, Hagit (2011) Compounds: the view from Hebrew. In: Lieber, R. & P. Štekauer (eds.) *The Oxford handbook of compounding*. Oxford: Oxford University Press.
- Borer, Hagit (2013a) *Taking Form. Structuring Sense III*. Oxford: Oxford University Press.
- Borer, Hagit (2013b) The syntactic domain of content. In Becker, M. et al. (eds.) *Generative linguistics and acquisition: Studies in honor of Nina M. Hyams*. Philadelphia: Benjamins.
- Botha, Rudolf P. (1968) *The function of the lexicon in transformational generative grammar. Janua Linguarum. Series Maior, Band 38*. The Hague: Mouton.
- Broekhuis, Hans (2015) *Syntax of Dutch: Verb and verb phrases. Volume 2*. Amsterdam: Amsterdam University Press.
- Bruening, Benjamin (2010) Ditransitive Asymmetries and a Theory of Idiom Formation. *Linguistic Inquiry* 41(4). 519-562.
- Bruening, Benjamin (2018) Word formation is syntactic: raising in nominalizations. *Glossa* 3(1). 102
- Brysbaert, Marc, & Boris New (2009) Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods* 41. 977-990. doi:10.3758/BRM.41.4.977
- Bye, Patrik (2007) Allomorphy: selection, not optimization. In: Sylvia Blaho, Patrik Bye & Martin Krämer (eds.) *Freedom of Analysis? (Studies in Generative Grammar)*, 63-92. Berlin: Mouton de Gruyter.
- Bye, Patrik & Peter Svenonius (2010) *Exponence, phonology and non-concatenative morphology*. University of Tromsø / CASTL. Ms. Lingbuzz/001099.

- Bye, Patrik & Peter Svenonius (2012) Non-concatenative morphology as epiphenomenon. In: Trommer, Jochen (ed.) *The morphology and phonology of exponence*, 427-495. Oxford: Oxford University Press.
- Cable, Seth (2004) *Phonological noun-verb dissimilarities in Optimal Paradigms*. Ms., MIT.
- Camacho, José (2012) Ser and estar: The individual/stage-level distinction and aspectual predication. In: José Ignacio Hualde et al. (eds.) *The Handbook of Hispanic Linguistics*, 453-475. Malden & Oxford: Blackwell Publishing.
- Capodieci, Frank, Bruce Hayes and Colin Wilson (2009) *Manual: Phonotactic Learning program*. University of California, Los Angeles, Manuscript.
- Cavirani-Pots, Cora (2020) *Roots in Progress. Semi-lexicality in the Dutch and Afrikaans verbal domain*. Utrecht: LOT Publications.
- Chomsky, Noam (1959) A review of B. F. Skinner's "Verbal behavior". *Language* 35.26-58.
- Chomsky, Noam (1965) *Aspects of the theory of syntax*. Cambridge, MA: The MIT Press.
- Chomsky, Noam (1970) Remarks on Nominalization. In: Jacobs, R. and P. Rosenbaum (eds.) *Readings in English Transformational Grammar*. Waltham, MA: Ginn.
- Chomsky, Noam (1995) *The Minimalist Program*. Cambridge, MA: The MIT Press.
- Chomsky, Noam (2001) Derivation by phase. In: Michael Kenstowicz (ed.) *Ken Hale: A Life in Language*, 1-52. Cambridge, MA: The MIT Press.
- Cinque, Guglielmo (1999) *Adverbs and functional heads. A cross-linguistic perspective*. Oxford Studies in Comparative Syntax. New York & Oxford: Oxford University Press.
- Cohen, Antonie (1958) Het Nederlandse diminutiefsuffix: Een morfonologische proeve. *De Nieuwe Taalgids* 51. 40-45.
- Corver, Norbert (1990) *The syntax of left branch constructions*. Doctoral dissertation, Tilburg University.
- Corver, Norbert (1997) The internal syntax of the Dutch extended adjectival projection. *NLLT* 15. 289-368.
- Corver, Norbert (2017) *Adverbial -s: so awks but so natural!* Ms., Utrecht University.
- De Belder, Marijke (2009) *On the syntax of titles*. Talk presented at ConSOLE, Nova Gorica, January 16-18, 2009.
- De Belder, Marijke (2011a) *Roots and affixes: Eliminating lexical categories from syntax*. LOT Dissertations. Utrecht: LOT.
- De Belder, Marijke (2011b) A morphosyntactic decomposition of countability. *The Journal of Comparative Germanic Linguistics* 14:3.
- De Belder, Marijke (2014) Review of Hagit Borer, Structuring Sense, vol. III: Taking Form. Oxford: Oxford University Press, 2013. Pp. xxvi + ^&!. *Journal of Linguistics* 50(3), pp. 718-725.
- De Belder, Marijke (2017) The root and nothing but the root: primary compounds in Dutch. *Syntax. A Journal of Theoretical, Experimental and Interdisciplinary Research* 20(2). 138-169.
- De Belder, Marijke (2020) A split approach to the selection of allomorphs: vowel length alternating allomorphy in Dutch. *Glossa* 5(1). 42. DOI: 10.5334/gjgl.899.
- De Belder, Marijke (to appear) The extravagant Dutch affix -ke and its meandering through the interfaces. In: Eitelmann, Matthias and Haumann, Dagmar (eds.) *Extravagant Morphology: Studies in rule-bending, pattern-extending and theory-challenging morphology. Studies in Language Companion Series*. Amsterdam and Philadelphia: John Benjamins Publishing Company.

- De Belder, Marijke (to appear) The syntax-semantics interface in Distributed Morphology. To appear in: A. Alexiadou, R. Kramer, A. Marantz & I. Oltra-Massuet (eds). *The Cambridge Handbook of Distributed Morphology*. Cambridge: Cambridge University Press.
- De Belder, Marijke, Noam Faust & Nicola Lampitelli (2014) On a low and a high diminutive: Evidence from Italian and Hebrew. In: Alexiadou, A. et al. (eds.) *The syntax of roots and the roots of syntax. Oxford studies in Theoretical Linguistics*. Oxford: Oxford University Press.
- De Belder, Marijke and Jeroen van Craenenbroeck (2013) *On vocabulary insertion*. HUBrussel, ms.
- De Belder, Marijke & Jeroen Van Craenenbroeck (2015) How to Merge a Root. *Linguistic Inquiry* 46(4). 625-655.
- De Belder, Marijke and Marjo van Koppen (2012) *On AN(N) compounds*. Talk presented at BCGL7, Brussels, December 17-18, 2012.
- De Belder, Marijke, Van Koppen, Marjo (2016). One module, different levels of merge: AN(N) compounds in Dutch. *Studia Linguistica*, 70(1). 1-33. DOI: 10.1111/stul.12040
- De Clercq, Karen & Guido Vanden Wyngaerd (2017) \*ABA Revisited: Evidence from Czech and Latin degree morphology. *Glossa: a journal of general linguistics* 2(1).
- De Haas, Wim & Mieke Trommelen (1993) *Morfologisch handboek van het Nederlands: een overzicht van de woordvorming*. 's-Gravenhage: SDU Uitgeverij.
- de Saussure, Ferdinand (1972, reprint of 1916) *Cours de linguistique Générale*. Paris: Payot.
- Den Dikken, Marcel (2006) *Relators and Linkers: The Syntax of Predication, Predicate Inversion, and Copulas*. *Linguistic Inquiry Monograph*. Cambridge, MA: The MIT Press.
- Den Dikken, Marcel (2008) *Directions from the GET-GO*. Keynote talk at the 10th Seoul International Conference on Generative Grammar. Seoul July 17-20, 2008.
- Delfitto, Denis, Antonio Fábregas and Chiara Melloni. (2008) *Compounding at the interfaces*. Talk presented at NELS 39, 2008.
- De Schutter, Georges, Boudewijn van den Berg, Ton Goeman & Thera de Jong (2005) *Morphological Atlas of the Dutch dialects. Volume 1*. Amsterdam: Amsterdam University Press.
- De Sutter Gert, Dirk Speelman, Dirk Geeraerts (2008) Prosodic and syntactic-pragmatic mechanisms of grammatical variation: the impact of postverbal constituent on the word order in Dutch clause final verb clusters. *International Journal of Corpus Linguistics* 13(2). 194-224.
- Diesing, Molly (1988) Bare Plural Subjects and the Stage/Individual Contrast. In: Manfred Krifka (ed.) *Generativity in Natural Language: Proceedings of the 1988 Tilbingen Conference*, 107–154. Tübingen: University of Tübingen.
- Di Sciullo, Anna Maria & Edwin Williams (1987) *On the definition of word*. Cambridge, MA.: The MIT Press.
- Don, Jan (1993) *Morphological Conversion*. Utrecht: OTS.
- Don, Jan (2009) IE, Germanic: Dutch. In Lieber, R. & P. Štekauer (eds.) *The Oxford handbook of compounding*. Oxford: Oxford University Press.
- Don, Jan & Marian Erkelens (2006) Vorm en Categorie. *Taal & Tongval* 19. 40-54.
- Doron, Edit (2003) Agency and voice: The semantics of semitic templates. *Natural Language Semantics* 11. 1–67.
- Doron, Edit (ed.) (2014) On the identity of roots. *Theoretical Linguistics* 40(3/4). 225-474.
- Dowty, David R. (1979) *Word Meaning and Montague Grammar: The Semantics of Verbs and Times in Generative Semantics and in Montague's PTQ*. Vol. 7. *Studies in Linguistics and Philosophy*. Dordrecht: Springer Science & Business Media.

- Dresher Bezalel Elan & Aditi Lahiri (1991) Metrical coherence in Germanic. *Linguistic Inquiry* 11. 47-73.
- Eisenberg, P. (2013) *Das Wort. Grundriss der deutschen Grammatik. 4. Auflage*. Stuttgart & Weimar: J.B. Metzler.
- Embick, David (1997) *Voice and the Interfaces of Syntax*. Doctoral dissertation, University of Pennsylvania.
- Embick, David (2004) On the structure of resultative participles in English, *Linguistic Inquiry* 35(3). 355-392.
- Embick, David (2007) Linearization and local dislocation: derivational mechanics and interactions. *Linguistic Analysis* 33(3/4). 303-336.
- Embick, David (2010) *Localism versus globalism in morphology and phonology*. *Linguistic Inquiry Monographs*. Cambridge, MA: The MIT Press.
- Embick, David (2020) The motivation for roots in Distributed Morphology. *Annual Review of Linguistics* 7(1). 1-20.
- Embick, David & Alec Marantz (2008) Architecture & blocking. *Linguistic Inquiry* 39(1). 1-53.
- Embick, David & Rolf Noyer (2001) Movement Operations after Syntax. *Linguistic Inquiry* 32(4). 555-595.
- Filip, Hana (1999) *Aspect, Eventuality Types, and Nominal Reference*. New York & London: Garland Publishing.
- Folli, Raffaella & Heidi Harley (2005) Flavours of v: Consuming Results in Italian and English. In: Paula Kempchinsky and Roumyana Slabakova (eds.) *Aspectual Enquiries*, 95–120. Dordrecht: Springer.
- Folli Raffaella, Christina Sevdali & Robert Truswell (eds.) (2013) *Syntax and its limits*. Oxford: Oxford University Press.
- Garrett, Susan (1996) Another look at Spanish stress and syllable structure. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 32. 61–75.
- Gervain, Judit & Janet F. Werker (2008) How infant speech perception contributes to language acquisition. *Language and Linguistics Compass* 2/6.
- Giorgi, Alessandra, & Fabio Pianesi (1997) *Tense and aspect: From semantics to morphosyntax*. Oxford: Oxford University Press.
- Goldsmith, John (1976) *Autosegmental Phonology*. PhD dissertation, MIT.
- Goodenough, Ward and Hiroshi Sugita (1980) *Trukese-English Dictionary*. Philadelphia: American Philosophical Society.
- Grimm, Scott (2012) *Number and Individuation*. PhD Dissertation, University of Stanford.
- Guimarães, Maximiliano (2000) *In defense of vacuous projections in Bare Phrase Structure*. University of Maryland Working Papers in Linguistics, vol.9: 90-115.
- Gussenhoven, Carlos & Haike Jacobs (1998) *Understanding Phonology*. London: Hodder Education.
- Haegeman, Liliane (1998) Gender and wordmarkers in West-Flemish. University of Geneva, ms.
- Haegeman, Liliane & Hill, Virginia (2013) The syntactization of discourse. In: Raffaella Folli, Christina Sevdali & Robert Truswell (eds.) *Syntax and Its Limits*, 370-390. Oxford: Oxford University Press.
- Haeseryn et al. (1997) *Algemene Nederlandse Spraakkunst*. Groningen: Martinus Nijhoff Uitgevers.
- Halle, Morris (1997) Distributed Morphology: Impoverishment and Fission. In: Bruening, Benjamin, Yoonjung Kang and Martha McGinnis (eds.) *MITWPL 30*. Cambridge, MA: MIT Working Papers in Linguistics.

- Halle, Morris & Alec Marantz (1993) Distributed Morphology and the pieces of inflection. In: Hale, Kenneth & Samuel Jay Keyser (eds.) *The view from building 20*, 111-176. Cambridge: The MIT Press.
- Halle, Morris & Alec Marantz (1994) Some key features of Distributed Morphology. In: Benjamin Bruening, Yoonjung Kang and Martha McGinnis (eds.) *MIT Working Papers in Linguistics 30: Papers at the interface*, 275-288. Cambridge, MA: MIT Working Papers in Linguistics 30.
- Hanssen, Esther (2012) *Linking elements in compounds: Regional variation in speech production and perception*. Doctoral dissertation, Radboud University Nijmegen.
- Harbour, Daniel (2008) *Morphosemantic Number: From Kiowa Noun Classes to UG Number Features*. Dordrecht: Springer.
- Harley, Heidi (2002) Possession and the double object construction. In: P. Pica and J. Rooryck (eds.) *The linguistic variation yearbook*, vol. 2, Amsterdam: John Benjamins, pp. 29 – 68.
- Harley, Heidi (2005) How do verbs get their names? In N. Erteschik-Shir & T. Rapoport (eds.). *The Syntax of Aspect* 42–64. Oxford: OUP.
- Harley, Heidi (2008) When is a syncretism more than a syncretism? Impoverishment, metasyncretism and underspecification. In: Harbour, Daniel, David Adger and Susana Béjar (eds.) *Phi Theory*. Oxford: OUP.
- Harley, Heidi (2008) On the causative construction. In: Shigeru Miyagawa and Mamoru Saito, (eds.) *Handbook of Japanese Linguistics*, 20-53. Oxford: The Oxford University Press.
- Harley, Heidi (2009) Compounding in DM. In Lieber, R. & P. Štekauer (eds.) *The Oxford handbook of compounding*, 129-144. Oxford: The Oxford University Press.
- Harley, Heidi (2012) Semantics in Distributed Morphology. In: C. Maienborn, K. von Heusinger and P. Portner (eds.) *Semantics: An International Handbook of Natural Language Meaning*, vol.3, 2151-2172. Berlin: Mouton de Gruyter.
- Harley, Heidi (2013) External arguments and the mirror principle: on the distinctness of Voice and v. *Lingua* 125. 34-57.
- Harley, Heidi (2014) On the Identity of Roots. *Theoretical Linguistics* 40(3/4). 255-276.. DOI: 10.1515/tl-2014-0010
- Harley, Heidi (2017) The 'bundling' hypothesis and the disparate functions of little v. In: d'Alessandro, R., I. Franco & Á. J. Gallego (eds.) *The Verbal Domain*, 255-278. Oxford: The Oxford University Press.
- Harley, Heidi & Rolf Noyer (1999) State-of-the-Article: Distributed Morphology. *Glot International* 4(4). 3-9.
- Harris, Alice C. (2017) *Multiple Exponence*. Oxford: The Oxford University Press.
- Harris, James W. (1983) *Syllable structure and stress in Spanish: A nonlinear analysis*. Cambridge, MA: The MIT Press.
- Harris, James W. (1991) The exponence of gender in Spanish. *Linguistic Inquiry* 22. 27-62.
- Hayes, Bruce & Colin Wilson (2007) A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39: 379-440.
- Heemskerk, Josée & Wim Zonneveld (2000) *Uitspraakwoordenboek*. Utrecht & Antwerpen: Unieboek/Het Spectrum.
- Heim, Irene (2008) Features on Bound Pronouns. In: Daniel Harbour, David Adger and Susana Béjar (eds.) *Phi Theory: Phi-features across Modules and Interfaces*. Oxford: The Oxford University Press.

- Heim, Johannes, Hermann Keupdjio, Zoe Wai-Man Lam, Adriana Osa- Gómez, Sonja Thoma & Martina Wiltschko (2016) Intonation and Particles as Speech Act Modifiers: A Syntactic Analysis. *Studies in Chinese Linguistics* 37(2). 109-129. DOI: 10.1515/scl-2016-0005
- Hermans, Ben & Gertjan Postma (2009) *The syntax of class*. Talk presented at the workshop Atoms and Laws of the Noun Phrase, Utrecht, 02-07-2009.
- Hill, Virginia (2013) *Vocatives: How Syntax meets with Pragmatics*. Leiden: Brill.
- Hoeksema, Jack (1988) Head-types in morpho-syntax., In: G. Booij and J.v. Marle (eds.) *Yearbook of Morphology 1988*, 123-38. Dordrecht: Foris Publications.
- Holmberg, Anders & Qi Wang (2019) Roots, categorizers and reduplication in Xining Chinese. In: Emonds, J. et al. (eds.) *Language Use and Linguistic Structure. Proceedings of the Olomouc Linguistic Colloquium 2018*. Olomouc: Palacký University Olomouc.
- Hünig, Matthias (2010) Adjective + Noun constructions between syntax and word formation in Dutch and German. In: Michel, Sascha and Alexander Onysko (eds.) *Cognitive approaches to word formation*. Berlin: De Gruyter Mouton. 195-218.
- Inkelas, Sharon (1998) The Theoretical Status of Morphologically Conditioned Phonology: A Case Study of Dominance Effects. In: Booij, G. & J. van Marle (eds.) *Yearbook of Morphology 1997*, Dordrecht: Kluwer Academic Publishers, pp. 121–55.
- Itô, Junko and Armin Mester (1995) Japanese Phonology. In: J. A. Goldsmith (ed.) *The Handbook of Phonological Theory*. Oxford: Blackwell, pp. 817–38.
- Josefsson, Gunlög (1997) *On the principles of word formation in Swedish*. PhD. Diss: Lund University.
- Kager, René (2008) Lexical irregularity and the typology of contrast. In: Kristin Hanson & Sharon Inkelas (eds.) *The nature of the word: Essays in honor of Paul Kiparsky*, 397-432. Cambridge, MA: The MIT Press.
- Kager, René & Wim Zonneveld (1985-1986) Schwa, syllables and extrametricality in Dutch. *The Linguistic Review* 5. 197-221.
- Kallulli, Dalina (2007) Rethinking the passive/anticausative distinction. *Linguistic Inquiry* 38(4). 770-780.
- Kastner, Itamar & Vera Zu (2017) Blocking and Paradigm Gaps. *Morphology* 27. 643-684.
- Kayne, Richard (2009) Antisymmetry and the Lexicon. In: Van Craenenbroeck, Jeroen (ed.) *The Linguistic Variation Yearbook 2008*, 1-32. Amsterdam: John Benjamins Publishing.
- Kayne, Richard (2016) *What is suppletive allomorphy? On ‘went’ and on “\*goed” in English*. Ms. New York University. Lingbuzz/003241.
- Kennedy, Christopher (1999) *Projecting the adjective: the syntax and semantics of gradability and comparison*. New York: Garland.
- Keuleers, Emmanuel, Marc Brysbaert & Boris New (2010) SUBTLEX-NL: A new frequency measure for Dutch words based on film subtitles. *Behavior Research Methods*, 42(3). 643-650.
- Kihm, Alain (2005) Noun class, gender, and the lexicon-syntax-morphology interfaces: a comparative study of Niger-Congo and Romance languages. In: Cinque, Guglielmo and Richard S. Kayne (eds.) *The Oxford Handbook of Comparative Syntax*, 459-512. Oxford: The Oxford University Press.
- Kiparsky, Paul (1982) From cyclic phonology to lexical phonology. In: Van der Hulst, H. & N. Smith (eds.) *The structure of phonological representations*, pt1, 130-175. Dordrecht: Foris Publications.
- Klockmann, Heidi (2017) *The design of semi-lexicality: evidence from case and agreement in the nominal domain*. LOT Dissertations. Utrecht: LOT.

- Kooij, Jan & Marc van Oostendorp (2003) *Fonologie. Uitnodiging tot de klankleer van het Nederlands*. Amsterdam: Amsterdam University Press.
- Koontz-Garboden, Andrew (2009) Anticausativization. *Natural Language and Linguistic Theory* 27. 77-138.
- Koopman, Hilda (2000) Prepositions, Postpositions, Circumpositions, and Particles. In Hilda Koopman (ed.) *The Syntax of Specifiers and Heads*, 204–60. London: Routledge.
- Koul, Omkar. sine dato. *Modes of Greetings in Kashmiri*. Ms., Indian Institute of Language Studies.
- Kracht, Marcus (2002) On the Semantics of Locatives. *Linguistics and Philosophy* 25. 157–232.
- Kramer, Ruth (2009) *Definite markers, phi features and agreement: A morphosyntactic investigation of the Amharic DP*. Santa Cruz, CA: University of California dissertation.
- Kramer, Ruth (2019) A novel kind of gender syncretism. In: Mathieu, Eric, Myriam Dali and Gita Zareikar (eds.) *Gender and Noun Classification*. Oxford: The Oxford University Press.
- Kratzer, Angelika (1995). Stage-level and individual-level predicates. In: G. N. Carlson and F. J. Pelletier (eds.) *The Generic Book*, 125–175. Chicago and London: University of Chicago Press.
- Kratzer, Angelika (1996) Severing the External Argument from Its Verb. In: Johan Rooryck and Laurie Zaring (eds.) *Phrase Structure and the Lexicon. Studies in Natural Language and Linguistic Theory* 33, 109–39. Dordrecht: Kluwer Academic Publishers.
- Kratzer, Angelika (2001) Building Statives. In *Proceedings of the Berkeley Linguistics Society* 26.
- Kratzer, Angelika (2013) *Creating a family: constructing transfer of possession verbs*. Paper presented at the little v Workshop in Leiden, October 2013.
- Krifka, Manfred (1998) The origins of telicity. In: Rothstein, S. (ed.) *Events and Grammar*, 197-235. Dordrecht: Kluwer.
- Krott, Andrea, Loes Krebbers, Robert Schreuder and R. Harald Baayen (2009) Semantic Influence on Linkers in Dutch Noun-Noun Compounds. *Folia Linguistica* 36(1-2). 7-22.
- Lahiri, Aditi & Bezalel Elan Dresher (1999) Open syllable lengthening in West Germanic. *Language* 75(4). 678-719.
- Lapointe Steven G. (1980) *A Theory of Grammatical Agreement*. Ph.D. diss., University of Massachusetts, Amherst.
- Leben, William (1973), *Suprasegmental Phonology*. PhD dissertation, MIT.
- Legendre, Geraldine, Yoshiro Miyata & Paul Smolensky (1990) Harmonic Grammar: A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. In: *Proceedings of the twelfth annual conference of the Cognitive Science Society*, 388–395. Cambridge, MA: Lawrence Erlbaum.
- Lemmens, Maarten (2002) The Semantic Network of Dutch Posture Verbs. In: Newman, John (ed.) *The Linguistics of Sitting, Standing and Lying, Vol. 51. Typological Studies in Language (TSL)*, 103-139. Amsterdam & Philadelphia: John Benjamins Publishing Company.
- Levelt, Willem J.M. (1999) Models of word production. *Trends in Cognitive Sciences* 3(6). 223-232.
- Levin, Beth (1993) *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press.
- Levin, Beth and Malka Rappaport Hovav (1995) *Unaccusativity: At the Syntax–Lexical Semantics Interface (Linguistic Inquiry Monograph 26)*. Cambridge, MA: The MIT Press.
- Levin, Beth, and Malka Rappaport Hovav (2005) *Argument Realisation*. Cambridge, MA: Cambridge University Press.

- Longobardi, Giuseppe (1994) Reference and proper names: a theory of N-movement in syntax and logical form. *Linguistic Inquiry* 25(4). 609-665.
- Lowenstamm, Jean (2007) On little n, ROOT, and types of nouns. In: Jutta Hartmann, Veronika Hegedus & Henk van Riemsdijk (eds.) *The sounds of silence*, 105-144. Amsterdam: Elsevier.
- Lowenstamm, Jean & Koos Van der Wilt (1982) *Dutch Diminutives: A Study in Metrical Ambiguity*. Ms., University of Texas at Austin.
- Lyons, Christopher (1999) *Definiteness*. Cambridge: Cambridge University Press.
- Maienborn, Claudia (2005) A discourse-based account of Spanish ser/estar. *Linguistics* 43(1). 155-180.
- Marantz, Alec (1984) *On the nature of grammatical relations*. Cambridge, MA: The MIT Press.
- Marantz, Alec (1997) No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. In Alexis Dimitriadis & Laura Siegel (eds.) *Proceedings of the 21st Annual Penn Linguistics Colloquium: Penn Working Papers in Linguistics* 4(2). 201-225.
- Marantz, Alec (2007) Phases and Words. In: Choe, S.D. (ed.) *Phases in the Theory of Grammar*, 191-222. Seoul: Dong-In Publishing Co.
- Marantz, Alec (2009) *Roots, re-, and affected agents: Can roots pull the agent under little v?* Talk given at Roots workshop, Universität Stuttgart, Stuttgart, Germany, June 10-13.
- Marantz, Alec (2013) Verbal argument structure: events and participants. *Lingua* 130. 152-168.
- Marelj, Marijana (2004) *Middles and argument structure across languages*. Utrecht: UIL-OTS.
- Marian, Viorica, James Bartolotti, Sarah Chabal, Anthony Shook (2012). CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. *PLoS ONE* 7(8): e43230. doi:10.1371/journal.pone.0043230
- Marušić, Franc (2005) *On non-simultaneous phases*. PhD Diss., Stony Brook University, Stony Brook NY.
- Marušić, Franc & Rok Žaucer (2008) Affix/Word Variation of Syntactically Parallel Structures and Distributed Morphology. Young-Sun Kim (ed.) *Minimalist Explorations of the Syntax-Lexicon Interface*, 263-282. Seoul: Hankuk Publishing Co.
- Mattens, Wim H.M. (1984) De voorspelbaarheid van tussenklanken in nominale samenstellingen. *De Nieuwe Taalgids* 77(4). 333-343.
- McCarthy, John J. and A. Prince (1993) *Prosodic Morphology I: Constraint Interaction and Satisfaction*, University of Massachusetts, Amherst, and Rutgers University.
- McCarthy, John J. & Alan Prince (1993) Generalized alignment. In: Geert Booij & Jaap van Marle (eds.) *Yearbook of Morphology 1993*, 79-153. Dordrecht: Kluwer.
- McCarthy, John J. (2005) Optimal paradigms. *Paradigms in Phonological Theory* 55.
- McCarthy, John J. (2007) *Hidden generalizations: Phonological opacity in Optimality Theory*. London: Equinox.
- McCawley, James D. (1968) *The phonological component of a grammar of Japanese*. The Hague & Paris: Mouton.
- McCready, Elin (2019) *The Semantics and Pragmatics of Honorification*. Oxford: Oxford University Press.
- McGinnis, Martha (2002) On the systematic aspect of idioms. *Linguistic Inquiry* 33(4). 665-672.
- Mithun, Marianne (1984) The Evolution of Noun Incorporation. *Language* 60(4). 847-894.
- Morgan, James L., & Katherine Demuth (eds.). (1996) *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

- Moshi, Lioba (1995) *Locatives in KiVunjo-Chaga*. In: Akinlabi, Akinbiyi (ed.) *Approaches to African linguistics*, 129-146. Trenton, NJ: Africa World Press, Inc.
- Muller, Jennifer (1999) A unified mora account of Chuukese. *Proceedings of the West Coast Conference on Formal Linguistics* 18. 393–405.
- Myler, Neil (2016) *Building and Interpreting Possession Structures*. Cambridge, MA: The MIT Press.
- Neijt, Anneke H., R. Harald Baayen and Robert Schreuder (2006) Formal and semantic constraints on the interpretation of the suffix s in reading Dutch nominal compounds. *Written Language and Literacy* 9(2). 247-264.
- Neijt, Anneke H. and Robert Schreuder (2009) De rol van en en s in samenstellingen. Veranderend Nederlands verandert gedachten over het taalsysteem. In Boogaert, R. et al. (eds.) *Woorden wisselen*, 93-104. Leiden: Stichting Nederlandstiek Leiden.
- Nevins, Andrew (2008) Cross-modular Parallels in the Study of Phon and Phi. In: Harbour, D. D. Adger and S. Béjar (eds.) *Phi Theory: Phi-features across Modules and Interfaces*. Oxford: OUP.
- Nevins, Andrew (2010) Phonologically-conditioned allomorph selection. In: Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.) *The Blackwell Companion to Phonology*, 2357-2382. Malden & Oxford: Blackwell Publishing Ltd.
- Nevins, Andrew (2011) Marked Targets versus Marked Triggers and Impoverishment of the Dual. *Linguistic Inquiry* 42(3). 413-444.
- New, Boris, Marc Brysbaert, Jean Veronis, & Christophe Pallier (2007) The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics* 28. 661-677.
- Newman, John (ed.) (2002) *The Linguistics of Sitting, Standing and Lying. Vol. 51. Typological Studies in Language (TSL)*. Amsterdam & Philadelphia: John Benjamins Publishing Company.
- Newmeyer, Frederick J. (2009) Current challenges to the lexicalist hypothesis: an overview and a critique. In: William D. Lewis, Simin Karimi, Heidi Harley and Scott O. Farrar (eds.) *Time and Again: Theoretical perspectives on formal linguistics in honor of D. Terence Langendoen*, 91-117. Amsterdam: John Benjamins.
- No author / no date. *Oudnederlands Woordenboek*. Leiden: Instituut voor Nederlandse Lexicologie.
- No author / no date. *Woordenboek der Nederlandsche Taal*. Leiden: Instituut voor Nederlandse Lexicologie.
- No author (2020) *Dikke Van Dale Online: groot woordenboek van de Nederlandse taal*. Utrecht and Antwerpen: Van Dale Uitgevers.
- Nóbrega, Vitor A. (2020) No escape from categorization: an insider's view on compounds. *Ilha do Desterro* 2020.
- Noyer, Rolf (1992). *Features, Positions and Affixes in Autonomous Morphological Structure*. PhD thesis, MIT.
- Noyer, Rolf (1997) *Features, Position and Affixes in Autonomous Morphological Structure*. New York: Garland Publishing.
- Nunberg, Geoffrey (1978) *The Pragmatics of Reference*. Bloomington: Indiana University Linguistics Club.
- Olson, Kenneth S. (2005) *The phonology of Mono*. Dallas: SIL International & University of Texas at Arlington.
- Orgun, C. O. (1996) *Sign-Based Morphology and Phonology with special attention to Optimality Theory*, Doctoral dissertation, Department of Linguistics, University of California,Berkeley.
- Partee, Barbara, Alice Ter Meulen & Robert Wall (1987) *Mathematical methods in linguistics*. Dordrecht: Kluwer Academic Publishers.

- Paster, Mary. 2006. *Phonological conditions on affixation*. Berkeley, CA: University of California dissertation.
- Pater, Joe (2009) Weighted Constraints in Generative Linguistics. *Cognitive Science* 33. 999-1035.
- Pierce, Marc & Hans C. Boas (2010) First Diminutive Formation and [d]-Epenthesis in Yiddish. *Interdisciplinary Journal for Germanic Linguistics and Semiotic Analysis* 15. 213-230.
- Postma, Gertjan (1990) *Morpho-phonemic compensatory lengthening in Frisian and Brazilian Portuguese*. Ms. Vrije Universiteit Amsterdam.
- Postma, Gertjan (2001) Negative polarity and the syntax of taboo. In: Hoeksema, Jack, Hotze Rullmann, Víctor Sánchez-Valencia and Ton van der Wouden (eds.) *Perspectives on Negation and Polarity Items*. *Linguistik Aktuell/Linguistics Today* 40, 283-330. Amsterdam & Philadelphia: John Benjamins Publishing Company.
- Prince, A. & P. Smolensky (1993) Optimality Theory: Constraint Interaction in Generative Grammar. *Rutgers University Center for Cognitive Science Technical Report* 2.
- Pustejovsky, James (1995) *The generative lexicon*. Cambridge, MA: The MIT Press.
- Pylkkanen, L. (2008) *Introducing Arguments*. Cambridge, MA: The MIT Press.
- Radford, Andrew (2000) NP shells. *Essex Research Reports in Linguistics* 33. 2-20.
- Ramchand, Gillian Catriona (2008) *Verb Meaning and the Lexicon: A First-Phase Syntax*. Cambridge: Cambridge University Press.
- Rappaport Hovav, Malka, and Beth Levin (1998) "Building Verb Meanings." In: Miriam Butt and Wilhelm Geuder (eds.) *The Projection of Arguments: Lexical and Compositional Factors*, 97-134. Stanford: CSLI Publications.
- Reinhart, T. (2002) The Theta System – An Overview. *Theoretical Linguistics* 28. 229–90.
- Ritter, Elizabeth (1991) Two functional categories in noun phrases: evidence from Modern Hebrew. In: Rothstein, S. (ed.) *Perspectives on phrase structure: heads and licensing*. San Diego: Academic Press.
- Roberts, Ian (1997) Restructuring, head movement, and locality. *Linguistic Inquiry* 28. 423-460.
- Ross, John R. (1970) On declarative sentences. In: Roderick A. Jacobs & Peter S. Rosenbaum (eds.) *Readings in English Transformational Grammar*, 222-272. Waltham, MA: Ginn & Co.
- Ruijsen, Marion (2011) *Oudnederlandse woordvorming. Een overzicht van Oudnederlandse afleidingen, samenstellingen, leenwoorden en leenvertalingen uit de Wachtendonckse Psalmen, vergeleken met hun Latijnse tegenhangers*. Universiteit Gent, master thesis.
- Sauerland, Uli (2008) On the Semantic Markedness of Phi-Features. In: Harbour, D. D. Adger and S. Béjar (eds.) *Phi Theory: Phi-features across Modules and Interfaces*. Oxford: OUP.
- Schäfer, Florian (2008) *The syntax of (anti-)causatives: External arguments in change-of-state contexts*. Amsterdam: John Benjamins Publishing Company.
- Schäfer, Florian (2009) The causative alternation. *Language and Linguistics Compass* 3(2). 641-681.
- Schäfer, Florian (2013). *Reconciling Burzio's generalization with unexpected accusatives: from German to Icelandic*. Talk presented at the 28th Comparative Germanic Syntax Workshop, Leipzig.
- Schoorlemmer, Erik (2009) *Agreement, dominance and doubling. The morphosyntax of DP*. Doctoral Dissertation, Leiden University.
- Schreuder, Robert, Anneke H. Neijt, Femke van der Weide & R. Harald Baayen (1998) Regular plurals in Dutch compounds: Linking graphemes or morphemes? *Journal of Language and Cognitive Processes*, 13:5 . 551-573.
- Selkirk, Elisabeth (1982) *The syntax of words*. Cambridge(MA): The MIT Press.

- Shi, Rushen, Janet F. Werker, & James Morgan (1999) Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition*, 72, B11–B21.
- Shi, Rushen, & Janet F. Werker (2003) Basis of preference for lexical words in six-month-old infants. *Developmental Science* 6. 484–488.
- Shi R., Janet F. Werker & Anne Cutler (2006) Recognition and representation of function words in English-learning infants. *Infancy* 10(2). 187–198.
- Siloni, Tal (1997) *Noun phrases and nominalisations*. Dordrecht: Kluwer Academic Publishers.
- Simpson, Andrew & Wei-tien Dylan Tsai (eds.), *Chinese Syntax in a Cross- Linguistic Perspective*, 429–441. Oxford & New York: Oxford University Press.
- Smith, J. L. (1997) Noun faithfulness: On the privileged status of nouns in phonology. *ROA* 242.
- Smith, J. L. (2011) Category-specific effects. In: Van Oostendorp M. et al. (eds.) *The Blackwell Companion to Phonology*. NY: John Wiley & Sons Ltd.
- Smolensky, P. & G. Legendre (2006) *The harmonic mind: from neural computation to optimality-theoretic grammar*. Cambridge, MA: The MIT Press.
- Starke, Michal (2009) Nanosyntax - a short primer to a new approach to language. *Nordlyd* 36:1. 1–6.
- Stassen, Leon (2009) *Predicative Possession*. Oxford: The Oxford University Press.
- Stiebels, Barbara (2013) Rule interaction in Kleverlandish Diminutive Formation. In: Fabian Heck & Anke Assmann (eds.) *Rule interaction in grammar (Linguistische Arbeitsberichte volume 90)*, 163–176. Leipzig: Universität Leipzig.
- Stump, Gregory T. (1989) A note on Breton pluralization and the Elsewhere Condition. *NLLT* 7. 261–273.
- Svenonius, Peter (2004) *Spatial prepositions in English*. Ms. University of Tromsø.  
<http://ling.auf.net/lingBuzz/000001>.
- Svenonius, Peter (2012) *Spanning*. Ms., University of Tromsø.
- Tang, Sze-Wing (2015) Cartographic syntax of pragmatic projections. In: Audrey Li, Simpson, Andrew & Wei-tien Dylan Tsai (eds.) *Chinese Syntax in a Cross- Linguistic Perspective*, 429-441. Oxford & New York: Oxford University Press.
- Thrainsson, H. (1996) On the (Non-)Universality of Functional Categories. Werner Abraham e.a. (eds.): *Minimal Ideas. Syntactic Studies in the Minimalist Framework*. Amsterdam: John Benjamins.
- Tiel, Bob van, Margit Rem and Anneke Neijt (2011) De historische ontwikkeling van de tussenklank in Nederlandse nominale samenstellingen. *Nederlandse Taalkunde* 16:2. 120–140.
- Tiersma, Pieter Meijes (1985) *Frisian reference grammar*. Dordrecht: Foris Publications.
- Travis, Lisa deMena (1984) *Parameters and effects of word order variation*. Doctoral dissertation, MIT.
- Trépos, Pierre (1957) *Le pluriel breton*. Brest: Emgleo Breiz.
- Trommelen, Mieke (1983) *The syllable in Dutch, with special reference to diminutive formation*. Dordrecht: Foris.
- Trommelen, M. (1989) Lettergreetstruktuur en woordkategorie. *De Nieuwe Taalgids* 82, 64-85.
- Van Craenenbroeck, J. (2014) On diagnosing complement-taking roots. *Theoretical Linguistics* 40(3/4). 361-374.
- Van der Hulst, Harry (1985) Ambisyllabicity in Dutch. In Hans Bennis & Frits Beukema (eds.). *Linguistics in the Netherlands* 1985, 57-67. Dordrecht: Foris.

- Van der Hulst, Harry (2008) The Dutch diminutive. In: C. Ewen, H. van der Hulst & N. Kula (eds.). *Trends in prosodic phonology*. Lingua volume 118:9. 1288-1306. DOI: 10.1016/j.lingua.2007.09.012
- Van den Toorn, M.C. (1975) Over de Semantische Kenmerken van Staan, Liggen En Zitten. *De Nieuwe Taalgids* 68:1. 459–64.
- Van Haeringen, Coenraad Bernardus (1947) De meervoudsvorming in het Nederlands. *Mededeelingen der Koninklijke Nederlandse Akademie van Wetenschappen. Nieuwe reeks, deel 10.* 131-156.
- Van Oostendorp, Marc. 1995. *Vowel quality and phonological projection*. Tilburg, The Netherlands: Tilburg University dissertation.
- Van Riemsdijk, Henk (1990) Functional Prepositions. In H. Pinkster and I. Gene (eds.) *Unity in Diversity*, 229–41. Dordrecht: Foris Publications.
- Van Riemsdijk, Henk, and Riny Huybregts (2002) Location and Locality. In Marc Van Oostendorp and Elena Anagnostopoulou (eds.) *Progress in Grammar: Articles at the 20th Anniversary of the Comparison of Grammatical Models Group in Tilburg*, 1-23. Amsterdam: Meertens Instituut.
- Van Zonneveld, Ron (1978) Verkleinwoordvorming. In: Ans Van Berkel, Willem Blok & Albert Sassen (eds.) *Proeven van Neerlandistiek. Aangeboden aan Prof. dr. Albert Sassen*, 279-302. Groningen: Nederlands Instituut.
- Venneman, Theo (1988) *Preference laws for syllable structure and the explanation of sound change: With special reference to German, Germanic, Italian, and Latin*. Berlin: Mouton de Gruyter.
- Veselinova, Ljuba N. (2006) *Suppletion in verb paradigms: Bits and pieces of the puzzle*. Amsterdam: John Benjamins Publishing Co.
- Williams, Edwin (1981) On the notions ‘lexically related’ and ‘head of a word’. *Linguistic Inquiry* 12. 245-274.
- Williams, Edwin (2003) *Representation Theory*. Cambridge, MA: The MIT Press.
- Williams, Edwin (2007) Dumping Lexicalism. In: Ramchand, Gillian and Charles Reiss (eds.) *The Oxford Handbook of Linguistic Interfaces*. Oxford: Oxford University Press.
- Wolf, Matthew Adam (2008) *Optimal interleaving: serial phonology-morphology interaction in a constraint-based model*. Amherst, MA: University of Massachusetts dissertation.
- Wood, Jim (2015) *Icelandic morphosyntax and argument structure*. Dordrecht: Springer.
- Wood, Jim, and Alec Marantz (2017) The Interpretation of External Arguments. In Roberta D'Alessandro, Irene Franco, and Ángel J. Gallego (eds.) *The Verbal Domain. Oxford Studies in Theoretical Linguistics*, 255-278. Oxford: Oxford University Press.
- Yu, Alan C.L. (2003) *The morphology and phonology of infixation*. Berkeley, CA: University of California dissertation.
- Zagona, Karen (2012) Ser and estar: Phrase structure and aspect. In: C. Nishida & C. Russi (eds.) *Building a Bridge between Linguistic Communities of the Old and the New World, Cahiers Chronos* 25. Amsterdam: Rodopi.
- Zec, Draga (2005) Prosodic differences among function words. *Phonology* 22. 77-112.
- Zonneveld, Wim (1976) *A formal theory of exceptions in generative phonology*. Dordrecht: De Gruyter.
- Zonneveld, Wim (1981-1982) The descriptive power of the Dutch theme-vowel. *Spektator* 11. 342-374.
- Zygis, Marzena (2010) Typology of consonantal insertions. *ZAS Papers in Linguistics* 52. 111 – 140